Australian Government

Patent Office
Canberra

I, MATTHEW WILLOUGHBY, ACTING TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. pr4701 for a patent by MURDOCH CHILDRENS RESEARCH INSTITUTE as filed on 02 May 2001.

COMMONWEALTH OF AUSTRALIA
AUSTRALIA
PATENT OFFICE

WITNESS my hand this
Fifth day of June 2009

MATTHEW WILLOUGHBY
ACTING TEAM LEADER
EXAMINATION SUPPORT AND SALES

Regulation 3.2

Murdoch Childrens Research Institute

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION

for the invention entitled:

"A molecular marker"

The invention is described in the following statement:

A MOLECULAR MARKER

FIELD OF THE INVENTION

The present invention relates generally to a molecular marker of the integrity of the extracellular matrix in an animal including a human subject. More particularly, the present invention provides a molecular marker of cartilage integrity. The identification of the molecular marker in circulatory or tissue fluid is indicative of disrepair of the extracellular matrix and in particular cartilage such as caused or facilitated by trauma or a degenerative disease or other condition, for example, arthritis or autoimmunity. The molecular marker is preferably in the form of a glycoprotein but the instant invention extends to genetic sequences encoding the polypeptide portion of the glycoprotein. Expression analysis of such genetic sequences provides predictive utility in detecting normal or abnormal extracellular matrix development. The identification of the molecular marker of the present invention enables the development of a range of diagnostic and therapeutic agents for degeneration of extracellular matrix or the poor development of the matrix at the fetal and postnatal stages. In a most preferred embodiment, the molecular marker is referred to herein as "WARP" for von Willebrand Factor A-Related Protein. The corresponding genetic form of WARP is referred to herein as *"WARP"*.

BACKGROUND OF THE INVENTION

Bibliographic details of the publications numerically referred to in this specification are collected at the end of the description.

Reference to any prior art in this specification is not, and should not be taken as, an acknowledgment or any form of suggestion that this prior art forms part of the common general knowledge in Australia or any other country.

The extracellular matrix (ECM) is a complex mixture of collagens, non-collagenous glycoproteins, and proteoglycans that interact to provide a structural scaffold as well as

specific cues for the maintenance, growth and differentiation of cells and tissues. The protein cores of a large number of ECM molecules are composed of different combinations of a finite collection of modules [1]. The conservation of amino acid sequence of these modules between different ECM proteins and protein families provides us with the opportunity to identify new proteins by database homology searching to help reveal additional modular ECM proteins.

One module present in a number of proteins is the type A-domain, first described in von Willebrand factor (reviewed in [2]). Members of the expanding von Willebrand factor type A-domain (VA) protein superfamily participate in a variety of functions including hemostasis, cell adhesion and protein-protein interactions between matrix molecules. ECM components that contain one or more VA domains include collagens types VI [3,4], VII [5], XII [6], XIV [7], and XX [8], matrilins-1, -2, -3, -4 (reviewed in [9]), cochlin [10], polydom [11] and nine transmembrane α integrin chains (α1, α2, α10, α11, αL, αM, αX, αD and αE) (reviewed in [12]) where they are also known as an 'I' domain. Non-matrix proteins that contain VA domains include complement system proteins (C2, B) [13], inter-α-trypsin inhibitor (subunits H1-H3) [14], $\alpha 2\beta$ subunit of L-type voltage-dependent Ca^{2+} channel [15], in addition to the archetypal VA domains of von Willebrand factor itself [16].

The crystal structure for several VA domains have been solved including the A1 [17] and A3 [18] domains of vWF, and the I domain of integrins αM [12], αL [19] and α2 [20]. These studies show that the VA module is an independently folding protein unit that attains a classic α/β 'Rossman' fold consisting of a parallel β sheet surrounded by amphipathic α helices, and in the majority of VA domains, a metal ion-dependent adhesion site (MIDAS) at the C-terminal end of the β sheet. The MIDAS motif which consists of five conserved amino acids (DxSxS, T, D) act together with surrounding residues to bind divalent cations and gives I domains of integrins their adhesive and ligand binding properties [12]. However, not all VA domains contain this motif, for example, the A1 and A3 A-domains of von Willebrand Factor lack some of these conserved amino acids and are

not predicted to bind metal ions [17,18] and the binding of collagen to the A3 domain is not metal ion dependent [18].

VA domains appear to play an important role in protein-protein interactions. In von Willebrand factor, they interact with subendothelial heparans, collagens I, III, (reviewed by [21]) and collagen VI [22]; in integrins the I domain interacts with several collagens [23]; and in collagen VI VA domains interact with heparin [24] and collagen IV [25]. In ECM molecules, the ability of VA domains to interact with other proteins and with each other to promote higher-order structure formation may be crucial in providing a linkage between ECM structural networks. For example, in collagen VI, a specific N-terminal α3(VI) collagen VA domain (N5) is important for the assembly of collagen VI tetramers into functional microfibrils [26] and in matrilin-1, interchain assembly and microfilament formation is promoted by the interaction of the VA domains in adjacent matrilin molecules [27].

In working leading up to the present invention, the inventors sought to further characterize the contribution of VA domain proteins to ECM structure and function. The inventors have now identified a new member of VA-domain protein superfamily referred to herein as von <u>W</u>illebrand factor <u>A</u> <u>R</u>elated-<u>P</u>rotein or WARP. WARP provides, therefore, a molecular marker of the integrity of the ECM and in particular cartilage. The inventors demonstrate that WARP is a novel disulfide-bonded oligomeric ECM glycoprotein that is expressed in cartilage. A genetic sequence encoding WARP is represented herein in itallicized form, i.e. *WARP*. Both WARP and *WARP* represent molecular markers of ECM and in particular cartilage integrity.

SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

Nucleotide and amino acid sequences are referred to by a sequence identifier number (SEQ ID NO:). The SEQ ID NOs: correspond numerically to the sequence identifiers <400>1, <400>2, etc. A sequence listing is provided after the claims.

The inventors have identified a molecular marker of ECM and in particular cartilage integrity in the form of a new member of the von Willebrand factor A (VA) domain superfamily of extracellular matrix proteins, which is referred to hercin as "WARP" for von Willebrand Factor A Related-Protein. To identify novel VA-containing proteins, the EST database at NCBI was searched using the N8 VA-type domain protein sequence from the α3(VI) collagen chain. A series of overlapping EST clones with homology to N8 that represented a novel VA protein was identified. The full-length WARP cDNA, referred to herein as *"WARP"*, is 2.3 kb in size and encodes a protein of 415 amino acids which contains, from the N-terminus, a putative signal sequence, a single VA-like domain, two fibronectin type III-like repeats, and a short proline and arginine-rich segment. Northern blot and Real-time (RT)-PCR analysis indicates that *WARP* is expressed highest in rib chondrocytes and MCT cells induced to express a hypertrophic chondrocyte-like phenotype. Using a polyclonal antibody raised against the VA domain, WARP was detected throughout all cartilage zones of the newborn tibial head by immunohistochemistry. In addition, WARP migrated as a disulfide-bonded oligomer in guanidine-soluble newborn mouse cartilage extracts by Western blotting. WARP, therefore, is a new member of VA domain superfamily of extracellular matrix proteins, which is expressed in cartilage and forms oligomers *in vivo*.

Accordingly, one aspect of the present invention provides an isolated polypeptide or a

derivative or homologue thereof which *in situ* forms part of the ECM in an animal wherein said polypeptide comprises a VA-related domain encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:1 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:1 or its complementary form under low stringency conditions.

Another aspect of the present invention provides an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in a mouse wherein said polypeptide comprises an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or its complementary form under low stringency conditions.

A further aspect of the present invention provides an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in a human wherein said polypeptide comprises an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:5 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:5 or its complementary form under low stringency conditions.

Still another aspect of the present invention contemplates an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in a mouse, said polypeptide comprising the amino acid sequence substantially as set forth in SE ID NO:4 or an amino acid sequence having at least about 65% similarity thereto.

Still a further aspect of the present invention provides an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in a human, said polypeptide comprising the amino acid sequence substantially as set forth in SE ID NO:6 or an amino acid sequence having at least about 65% similarity thereto.

Even still another aspect of the present invention provides an isolated nucleic acid

molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a polypeptide which *in situ* forms part of the ECM in an animal wherein said nucleotide sequence comprises a sequence substantially as set forth in SEQ ID NO:1 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:1 or its complementary form under low stringency conditions.

Even still a further aspect of present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a murine WARP or a derivative or homologue thereof, said nucleotide sequence substantially as set forth in SEQ ID NO:3 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or its complementary form under low stringency conditions.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a human WARP or a derivative or homologue thereof, said nucleotide sequence substantially as set forth in SEQ ID NO:5 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:5 or its complementary form under low stringency conditions.

Still yet another of the present invention provides a method for producing a recombinant WARP, said method comprising introducing a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms or a nucleotide sequence having at least about 65% similarity to SQ ID NO:3 or SEQ ID NO:5 or their complementary forms or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms under low stringency conditions into a cell, culturing the cell or population of cells under conditions sufficient to permit expression of said nucleic acid molecule and then recovering the recombinant polypeptide.

Even yet another aspect of the present invention provides a method of identifying a nucleotide sequence likely to encode a WARP, said method comprising interrogating an animal genome database conceptually translated into different reading frames with an amino acid sequence defining a VA domain and identifying a nucleotide sequence corresponding to a sequence encoding said VA domain.

Even still another aspect of the present invention contemplates a method of detecting a loss of ECM integrity in an animal subject, said method comprising screening body fluid from said animal for the presence of a WARP or fragment thereof wherein the presence of said WARP or fragment is indicative of a loss of ECM integrity.

Another aspect of the present invention contemplates, therefore, a method for detecting a WARP or fragment thereof in a biological sample from a subject, said method comprising contacting said biological sample with an antibody specific for said WARP or fragment thereof or its derivatives or homologues for a time and under conditions sufficient for an antibody-polypeptide complex to form, and then detecting said complex.

A further aspect of the present invention provides a cartilage-specific promoter or functional derivative or homologue thereof, said promoter *in situ* operably linked to a nucleotide sequence comprising SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms or a nucleotide sequence having at least about 65% similarity to SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms under low stringency conditions.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a representation of the structure and modular organization of WARP. **(A)** Nucleotide and deduced amino acid sequence of WARP. The stop codon at nucleotides 1275-1277 is marked with an asterix and a potential polyadenylation site at nucleotides 2279-2285 is shown in bold type. The position of potential N-linked (Asn^{264} and Asn^{359}) and O-linked (Ser^{148}, Thr^{361} and Thr^{400}) glycosylation sites are underlined. C-terminal cysteine residues (Cys^{369} and Cys^{393}) available for disulfide bond formation are circled. **(B)** The modular structure of WARP is shown using standard symbols to represent conserved ECM protein modules [51].

Figure 2 is a representation of the alignment of VA domain and F3 repeats of WARP with homologous domains in other ECM proteins. Identical positions are shown within dark boxes and conserved substitutions in grey boxes. Alignments were performed using CLUSTALW (http://www.ch.embnet.org/software/ClustalW.html) [52] and shaded using BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html). **(A)** Alignment of VA domains from several ECM and non-ECM proteins. Sequences are matrilin-2 (GenBank Accession # NP_058042, amino acids 55-239), matrilin-4 (NP_038620, 34-218), matrilin-3 (NP_034900, 76-260), matrilin-1 (NP_034899, 43-227), collagen XIV (S78476, 156-337), collagen XII (NP_004361, 2321-2503), collagen VII (NP_000085, 36-218), collagen VI (AAD01978, 36-219), WARP (32-212), cochlin (O42163, 160-142), VLA-1 α-integrin (P56199, 142-334), and A1 domain of vWF (NP_000543, 1275-1460). The asterix indicates the conserved residues within the metal-ion dependent adhesion site [12]. The species of the sequences indicated in parentheses are: m, mouse; h, human; ch, chicken. **(B)** Alignment of F3 repeats from a sample of ECM proteins. The β-strands are designated by letters A-G above alignment according to [40]. Sequences are WARP F3 domain 2 (308-394), collagen VII (NP_000085, 235-325), collagen XIV (S78476, 627-711), β4 integrin chain (NP_000204, 1461-1548), collagen XII (NP_004361, 726-810), fibronectin (P11276, 1635-1720), WARP F3 domain 1 (215-301) and tenascin R (1589549, 867-951).

Figure 3 is a representation showing that WARP is a secreted glycoprotein. *WARP*/His cDNA in pCEP4 was transfected into 293-EBNA human embryonic kidney cells and *WARP*/His protein was immunoprecipitated from cell layer (lanes 1 and 3) and medium (lanes 2, 4-5) fractions of untransfected 293-EBNA cells (293-EBNA, lanes 1 and 2) or 293-EBNA cells transfected with *WARP*/His cDNA (*WARP* 293-EBNA, lanes 3-5) using an anti-His antibody. Sample digested with N-Glycosidase F following immunoprecipitation is shown in lane 5. All samples were reduced with 20 mM DTT prior to SDS-PAGE. The migration position of molecular weight markers is indicated on the left.

Figure 4 is a photographic representation showing expression of *WARP mRNA* in mouse tissues and cell lines. **(A)** Northern blot analysis of *WARP*. Poly(A) mRNA isolated from primary mouse chondrocytes (lane 1), MC3T3 osteoblasts (lane 2), Mov13 fibroblasts (lane 3) and C2C12 myoblasts (lane 4) was fractionated on a 1% v/v agarose gel and transferred to nylon membrane. The membrane was probed with $\alpha[^{32}P]$dCTP-labeled *WARP* cDNA fragment and exposed to X-ray film. The migration position of RNA markers in kb is indicated on left. **(B)** RT-PCR analysis of *WARP mRNA* expression. Total RNA was isolated from mouse tissues (lanes 1-6) and cell lines (lanes 7-11), treated with DNase to remove contaminating genomic DNA, and added to an oligo d(T)-primed RT reaction followed by PCR using primers specific for *WARP* (upper panel) and HPRT (lower panel). **(C)** Real-time PCR analysis of *WARP mRNA* expression. Each reaction contained oligo d(T)-primed cDNA, primers and fluorescently-labeled probes specific for *WARP* and HPRT. Data are represented as *WARP* signal relative to HPRT signal. The cDNA templates used were: 1, primary rib chondrocytes; 2, de-differentiated chondrocytes; 3, MCT cells induced to a hypertrophic chondrocyte-like phenotype; 4, MCT cells induced to an osteoblast-like phenotype; 5, MCT chondrocytes induced to change from hypertrophic chondrocyte-like to osteoblast-like phenotype; 6, MC3T3 osteoblasts; 7, Mov13 fibroblasts; and 8, 3T3 fibroblasts.

Figure 5 is a photographic representation showing expression of WARP protein in mouse cartilage. **(A)** Western blot showing WARP expression in sequential joint cartilage

extracts. Lane 1, 170 ng of GST-VA domain fusion protein; lane 2, F1 extract containing material soluble in Tris/EDTA; lane 3, F2 extract containing material soluble following chondroitinase and hyaluronidase digestion of insoluble material remaining from F1 extract; lane 4, F3 extract is material soluble in 6 M guanidine derived from insoluble material following F2 extraction. The WARP antibody (1 in 1000 dilution) was used to probe the blot containing lanes 1-4. Lane 5, F3 extract probed with matrilin-1 antibody (1 in 500 dilution). Lanes 2-5 each contained 20 μg protein per lane and samples were reduced with 2 mM tributylphosphine and 2.5% v/v β-mercapto-ethanol prior to electrophoresis. The migration position of molecular weight markers is indicated on left. **(B)** WARP protein expression in cartilage. 10 μM sagittal sections of anterior tibia from newborn mice stained with WARP antisera. Left panel, section showing developing cartilage, and surrounding connective tissue. Right panel, higher magnification of boxed region showing hypertrophic and pre-proliferative zones.

Figure 6 is a photographic representation showing that WARP forms higher-order structures. Western blot showing WARP expression in guanidine-soluble extracts of newborn mouse rib cartilage. Lane 1, rib cartilage sample reduced with 2 mM tributylphosphine and 2.5% v/v β-mercapto-ethanol; lane 2, cartilage sample prepared and resolved in the absence of reducing agents; lane 3, 170 ng of GST-VA domain fusion protein. Lanes 1 and 2 contained 20 μg of protein per lane. WARP antibody used at a dilution of 1 in 1000. The migration position of the molecular weight markers is indicated on left.

A summary of sequence identifiers is provided below:-

SUMMARY OF SEQENCE IDENTIFIERS

SEQ ID NO:	DESCRIPTION
1	Nucleotide sequence of human VA domain
2	Amino acid sequence of human VA domain
3	Nucleotide sequence of mouse WARP
4	Amino acid sequence of mouse WARP
5	Nucleotide sequence of human WARP
6	Amino acid sequence of human WARP
7	Nucleotide sequence of mouse VA domain
8	Amino acid sequence of human VA domain
9	NR1 primer
10	NF4 primer
11	mHPRT1 primer
12	mHPRT2 primer
13	WARP probe
14	WARP primer
15	WARP primer
16	HPRT probe
17	HPRT primer
18	HPRT primer
19	genomic sequence of human WARP

A summary of the abbreviations used is provided below:-

ABBREVIATIONS

ABBREVIATION	DESCRIPTION
ECM	extracellular matrix
WARP	von Willebrand Factor A domain related-protein
WARP	genetic sequence encoding WARP
VA	von Willebrand Factor A domain
N-terminus	amino-terminus
C-terminus	carboxyl-terminus
EST	expressed sequence tag
FACIT	Fibril-Associated Collagens with Interrupted Triple-Helices
PCR	polymerase chain reaction
bp	base pairs
kDa	kilodalton
SDS	sodium dodecyl sulfate

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is predicated in part on the identification of a new member of the von Willebrand Factor A (VA) domain superfamily of extracellular matrix (ECM) proteins and to a genetic sequence encoding same. The novel polypeptide of the present invention and its encoding genetic sequence as well as derivatives, homologues and analogues thereof are useful as molecular markers of the integrity of the ECM and in particular cartilage and as indicators of disease, trauma or poor development in animal including human subjects. The instant polypeptide is referred to herein as "WARP" for von Willebrand Factor A-Related-Protein.

Accordingly, one aspect of the present invention provides an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in an animal wherein said polypeptide comprises a VA-related domain encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:1 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:1 or its complementary form under low stringency conditions.

The nucleotide sequence set forth in SEQ ID NO:1 represents the nucleotide sequence of the human VA domain. An example of a homologue of this sequence from a murine source is set forth in SEQ ID NO:7.

Reference herein to a "polypeptide" or a "WARP" or a protein form of a molecular marker includes a protein in a monomeric or oligomeric state and/or in a folded or unfolded state as well as a polypeptide associated with non-proteinaceous moieties such as carbohydrates, lipids or phosphate groups. Most preferably, the polypeptide is a glycoprotein. The term "glycoprotein" means a polypeptide associated with carbohydrate moieties as well as a glycosylated polypeptide. It is not the intention of the present invention to be limited solely to a glycoprotein since the polypeptide portion may have utility on its own such as its ability to induce antibody formation, in diagnostic assays and for therapeutic applications.

Reference herein to an "animal" includes any vertebrate animal comprising an ECM and in particular cartilage and includes humans, primates, livestock animals (e.g. sheep, goats, cows, pigs, horses, donkeys), companion animals (e.g. dogs, cats), laboratory test animals (e.g. mice, rats, rabbits, guinea pigs) and captured wild animals.

In one particularly preferred embodiment, the subject WARP is of murine origin and in particular mouse origin and comprises an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3.

Accordingly, another aspect of the present invention provides an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in a mouse wherein said polypeptide comprises an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or its complementary form under low stringency conditions.

In another embodiment, the instant polypeptide is of human origin and is encoded by a nucleic acid molecule substantially as set forth in SEQ ID NO:5. Such a polypeptide is referred to herein as human WARP.

According to this embodiment, there is provided an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in a human wherein said polypeptide comprises an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:5 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:5 or its complementary form under low stringency conditions.

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or

conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. In a particularly preferred embodiment, nucleotide and sequence comparisons are made at the level of identity rather than similarity.

Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include "reference sequence", "comparison window", "sequence similarity", "sequence identity", "percentage of sequence similarity", "percentage of sequence identity", "substantially similar" and "substantial identity". A "reference sequence" is at least 12 but frequently 15 to 18 and often at least 25 or above, such as 30 monomer units, inclusive of nucleotides and amino acid residues, in length. Because two polynucleotides may each comprise (1) a sequence (i.e. only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window" refers to a conceptual segment of typically 12 contiguous residues that is compared to a reference sequence. The comparison window may comprise additions or deletions (i.e. gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerized implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (i.e. resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as, for example, disclosed by Altschul *et al.* (1997) [53]. A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.* (1998) [54].

The terms "sequence similarity" and "sequence identity" as used herein refers to the extent that sequences are identical or functionally or structurally similar on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a "percentage of sequence identity", for example, is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g. A, T, C, G, I) or the identical amino acid residue (e.g. Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. For the purposes of the present invention, "sequence identity" will be understood to mean the "match percentage" calculated by the DNASIS computer program (Version 2.5 for windows; available from Hitachi Software engineering Co., Ltd., South San Francisco, California, USA) using standard defaults as used in the reference manual accompanying the software. Similar comments apply in relation to sequence similarity.

Preferably, the percentage (%) similarity or identity is at least about 70%, more preferably at least about 75%, still more preferably at least about 80%, even more preferably at least about 85%, yet even more preferably at least about 90-100% such as 91% or 92% or 93% or 94% or 95% or 96% or 97% or 98% or 99%.

Reference herein to a low stringency includes and encompasses from at least about 0 to at least about 15% v/v formamide and from at least about 1 M to at least about 2 M salt for hybridization, and at least about 1 M to at least about 2 M salt for washing conditions. Generally, low stringency is at from about 25-30°C to about 42°C. The temperature may be altered and higher temperatures used to replace formamide and/or to give alternative stringency conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5 M to at least about 0.9 M salt for hybridization, and at least about 0.5 M to at least about 0.9 M salt for washing

conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01 M to at least about 0.15 M salt for hybridization, and at least about 0.01 M to at least about 0.15 M salt for washing conditions. In general, washing is carried out $T_m = 69.3 + 0.41$ (G+C)% [55]. However, the T_m of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatch base pairs [56]. Formamide is optional in these hybridization conditions. Accordingly, particularly preferred levels of stringency are defined as follows: low stringency is 6 x SSC buffer, 0.1% w/v SDS at 25-42°C; a moderate stringency is 2 x SSC buffer, 0.1% w/v SDS at a temperature in the range 20°C to 65°C; high stringency is 0.1 x SSC buffer, 0.1% w/v SDS at a temperature of at least 65°C.

In a particularly preferred embodiment, the present invention is directed to an isolated polypeptide of human origin comprising a sequence of amino acids defining a VA-related domain and having an amino acid sequence substantially as set forth in SEQ ID NO:2 or an amino acid sequence having at least about 65% similarity thereto. A homologue of murine origin comprises a VA-related domain having the amino acid sequence set forth in SEQ ID NO:8.

Even more particularly, another aspect of the present invention contemplates an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in a mouse, said polypeptide comprising the amino acid sequence substantially as set forth in SE ID NO:4 or an amino acid sequence having at least about 65% similarity thereto.

In another embodiment, the present invention provides an isolated polypeptide or a derivative or homologue thereof which *in situ* forms part of the ECM in a human, said polypeptide comprising the amino acid sequence substantially as set forth in SE ID NO:6 or an amino acid sequence having at least about 65% similarity thereto.

As stated above, the polypeptide of the present invention is also referred to as "WARP" meaning a von Willebrand Factor A Related-Protein. Reference herein to a subject polypeptide or WARP includes reference to a derivative, homologue or analogue thereof.

The instant polypeptide or WARP is also referred to as a molecular marker.

A "derivative" includes a mutant, fragment, part, portion or hybrid molecule. A derivative generally but not exclusively carries a single or multiple amino acid substitution, addition and/or deletion.

A "homologue" includes an analogous polypeptide having at least about 65% similar amino acid sequence from another animal species or from a different locus within the same species.

Generally, the term "analogous polypeptide" means that the polypeptide or WARP is performing the same function or is part of the same structure between or within animal species. However, the present invention extends to any ECM protein including polypeptide having an amino acid sequence substantially at least about 65% similar to SEQ ID NO:4 or SEQ ID NO:6.

An "analogue" is generally a chemical analogue. Chemical analogues of the subject polypeptide contemplated herein include, but are not limited to, modification to side chains, incorporation of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with $NaBH_4$.

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitization, for example, to a corresponding amide.

Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acid, contemplated herein is shown in Table 1.

TABLE 1

Non-conventional amino acid	Code	Non-conventional amino acid	Code
α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
aminocyclopropane-	Cpro	L-N-methylasparagine	Nmasn
carboxylate		L-N-methylaspartic acid	Nmasp
aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl-	Norb	L-N-methylglutamine	Nmgln
carboxylate		L-N-methylglutamic acid	Nmglu
cyclohexylalanine	Chexa	L-Nmethylhistidine	Nmhis
cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
D-threonine	Dthr	L-norleucine	Nle
D-tryptophan	Dtrp	L-norvaline	Nva

D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib
D-valine	Dval	α-methyl-γ-aminobutyrate	Mgabu
D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
D-α-methylarginine	Dmarg	α-methylcylcopentylalanine	Mcpen
D-α-methylasparagine	Dmasn	α-methyl-α-napthylalanine	Manap
D-α-methylaspartate	Dmasp	α-methylpenicillamine	Mpen
D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
D-α-methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
D-α-methylisoleucine	Dmile	N-amino-α-methylbutyrate	Nmaabu
D-α-methylleucine	Dmleu	α-napthylalanine	Anap
D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
D-α-methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
D-α-methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
D-α-methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
D-α-methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
D-α-methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
D-α-methylvaline	Dmval	N-cylcododecylglycine	Ncdod
D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
D-N-methylhistidine	Dnmhis	N-(hydroxyethyl))glycine	Nser
D-N-methylisoleucine	Dnmile	N-(imidazolylethyl))glycine	Nhis

D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp
D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
γ-aminobutyric acid	Gabu	N-(*p*-hydroxyphenyl)glycine	Nhtyr
L-*t*-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
L-ethylglycine	Etg	penicillamine	Pen
L-homophenylalanine	Hphe	L-α-methylalanine	Mala
L-α-methylarginine	Marg	L-α-methylasparagine	Masn
L-α-methylaspartate	Masp	L-α-methyl-*t*-butylglycine	Mtbug
L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
L-α-methylleucine	Mleu	L-α-methyllysine	Mlys
L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
L-α-methylserine	Mser	L-α-methylthreonine	Mthr
L-α-methyltryptophan	Mtrp	L-α-methyltyrosine	Mtyr
L-α-methylvaline	Mval	L-N-methylhomophenylalanine	Nmhphe
N-(N-(2,2-diphenylethyl) carbamylmethyl)glycine	Nnbhm	N-(N-(3,3-diphenylpropyl) carbamylmethyl)glycine	Nnbhe

1-carboxy-1-(2,2-diphenyl-ethylamino)cyclopropane	Nmbc

Crosslinkers can be used, for example, to stabilize 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with n=1 to n=6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N $_\alpha$-methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

The present invention further contemplates chemical analogues of the subject polypeptide capable of acting as antagonists or agonists of the WARP or which can act as functional analogues of the WARP. Chemical analogues may not necessarily be derived from the instant polypeptide but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to mimic certain physiochemical properties of the subject polypeptide. Chemical analogues may be chemically synthesized or may be detected following, for example, natural product screening. The latter refers to molecules identified from various environmental sources such a river beds, coral, plants, microorganisms and insects.

These types of modifications may be important to stabilize the subject polypeptide if administered to an individual or for use as a diagnostic reagent.

Other derivatives contemplated by the present invention include a range of glycosylation variants from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in

different host cells.

The present invention further contemplates genetic sequences encoding the subject WARP. Such genetic sequences are referred to herein as *WARP*.

According to this embodiment, there is provided an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a polypeptide which *in situ* forms part of the ECM in an animal wherein said nucleotide sequence comprises a sequence substantially as set forth in SEQ ID NO:1 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:1 or its complementary form under low stringency conditions.

Another example of a nucleotide sequence encompassed by the above is the nucleotide sequence substantially set forth in SEQ ID NO:7.

In one preferred embodiment, the nucleic acid molecule is a murine *WARP* such as the nucleic acid molecule defined by SEQ ID NO:3.

In another embodiment, the nucleic acid molecule is a human *WARP* such as the nucleic acid molecule defined by SEQ ID NO:5.

Accordingly, another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a murine WARP or a derivative or homologue thereof, said nucleotide sequence substantially as set forth in SEQ ID NO:3 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or its complementary form under low stringency conditions.

In another embodiment, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence

encoding a human WARP or a derivative or homologue thereof, said nucleotide sequence substantially as set forth in SEQ ID NO:5 or its complementary form or a nucleotide sequence having at least about 65% similarity thereto or a nucleotide sequence capable of hybridizing to SEQ ID NO:5 or its complementary form under low stringency conditions.

The subject nucleic acid molecule may be DNA (e.g. cDNA or genomic DNA) or RNA (e.g. mRNA) or be an RNA:DNA hybrid. Furthermore, the nucleic acid molecule may have nucleotide analogues inserted to facilitate resistance, for example, to nucleases. The nucleotide sequence of the genomic clone of human WARP is represented in SEQ ID NO:19 and is encompassed by the invention. The cDNA sequence encoding WARP and its corresponding amino acid sequence are represented in SEQ ID NOS:5 and 6, respectively.

The nucleic acid molecule may be linear, single or double stranded or in a covalently closed, circular form.

In a particularly useful embodiment, the nucleic acid molecule is in a vector or plasmid such as but not limited to an expression vector. The use of vectors is a particularly convenient means of producing recombinant forms of the subject WARP.

According to this embodiment, there is provided a method for producing a recombinant WARP, said method comprising introducing a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms or a nucleotide sequence having at least about 65% similarity to SQ ID NO:3 or SEQ ID NO:5 or their complementary forms or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms under low stringency conditions into a cell, culturing the cell or population of cells under conditions sufficient to permit expression of said nucleic acid molecule and then recovering the recombinant polypeptide.

This aspect of the present invention extends to derivatives and homologues of the subject nucleic acid molecules such as nucleic acid molecules encoding functional portions of the

instant WARP. One example of a functional portion is a portion capable of interacting with another polypeptide or protein.

Although the present invention is particularly exemplified in relation to nucleic acid molecules defined by SEQ ID NO:3 or SEQ ID NO:5, the present invention extends to other related nucleic acid molecules which encode WARPs in the ECM. Such nucleic acid molecules are conveniently located by homology searching of particular databases.

According to this embodiment, there is provided a method of identifying a nucleotide sequence likely to encode a WARP, said method comprising interrogating an animal genome database conceptually translated into different reading frames with an amino acid sequence defining a VA domain and identifying a nucleotide sequence corresponding to a sequence encoding said VA domain.

Preferably, the genome is conceptually translated into from about 3 to about 6 reading frames and more preferably 6 reading frames.

The VA domain amino acid sequence may come from any convenient source such as but not limited to the 200 amino acid sequence of the $\alpha3$(VI) N8 VA domain of human collagen VI. Interrogation also may be by any convenient means such as using the tblastn (v2.0) program.

Alternatively, hybridization may be used to interrogate genomic or cDNA clones to identify related nucleotide sequences.

WARPs and their genetic sequences have a range of therapeutic and diagnostic utilities. For example, any compromise in the integrity of the ECM may result in WARP or fragments thereof being detected in circulatory or tissue fluid such as blood, urine, synovial or lymph fluid. The detection of a WARP or fragments thereof would be indicative of a degenerative or disease condition, trauma or infection. Examples of various conditions include autoimmune disease, arthritis, sporting injuries, osteoporosis and

various bone disorders. The detection of WARP in ECM and in particular cartilage is also indicative of normal ECM development. Accordingly, subjects may be tested *in utero* or post-natally for the presence of the WARP in the ECM to determine that ECM is developing correctly and is maintaining its integrity. Detection of the WARP in the ECM is also a useful monitor of regeneration of ECM following, for example, trauma or disease.

Accordingly, another aspect of the present invention contemplates a method of detecting a loss of ECM integrity in an animal subject, said method comprising screening body fluid from said animal for the presence of a WARP or fragment thereof wherein the presence of said WARP or fragment is indicative of a loss of ECM integrity.

Any number of detection methods may be employed. Immunological testing, however, is particularly convenient. Accordingly, the present invention extends to antibodies and other immunological agents directed to or preferably specific for said WARP or a fragment thereof. The antibodies may be monoclonal or polyclonal or may comprise Fab fragments or synthetic forms.

Specific antibodies can be used to screen for the subject WARP and/or their fragments. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies referred to above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of the WARP.

Both polyclonal and monoclonal antibodies are obtainable by immunization with a WARP or antigenic fragments thereof and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred

but are relatively easily prepared by injection of a suitable laboratory animal with an effective amount of subject polypeptide, or antigenic parts thereof, collecting serum from the animal and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates, therefore, a method for detecting a WARP or fragment thereof in a biological sample from a subject, said method comprising contacting said biological sample with an antibody specific for said WARP or fragment thereof or its derivatives or homologues for a time and under conditions sufficient for an antibody-polypeptide complex to form, and then detecting said complex.

The presence of the instant WARP or its fragment may be accomplished in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to U.S. Patent Nos. 4,016,043, 4, 424,279 and 4,018,653.

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule

capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain a subject polypeptide including by tissue biopsy, blood, synovial fluid and/or lymph. The sample is, therefore, generally a biological sample comprising biological fluid.

In the typical forward sandwich assay, a first antibody having specificity for the instant polypeptide or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or where more convenient, overnight) and under suitable conditions (e.g. for about 20°C to about 40°C) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be

labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules. In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated

by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescene and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

The present invention also contemplates genetic assays such as involving PCR analysis to detect RNA expression products of a genetic sequence encoding a WARP. Alternative methods or methods which may be used in conjunction include direct nucleotide sequencing or mutation scanning such as single stranded conformation polymorphoms analysis (SSCP) as well as specific oligonucleotide hybridization.

The present invention further contemplates kits to facilitate the rapid detection of WARPs or their fragments in a subject's biological fluid.

Still yet another aspect of the present invention contemplates genomic sequences including gene sequences encoding a WARP as well as regulatory regions such as promoters, terminators and transcription/translation enhancer regions associated with the gene encoding a WARP.

The term "gene" is used in its broadest sense and includes cDNA corresponding to the exons of a gene. Accordingly, reference herein to a 'gene' is to be taken to include:-

(i) a classical genomic gene consisting of transcriptional and/or translational regulatory sequences and/or a coding region and/or non-translated sequences (i.e. introns, 5'- and 3'- untranslated sequences); or

(ii) mRNA or cDNA corresponding to the coding regions (i.e. exons) and 5'- and 3'-untranslated sequences of the gene.

The term "gene" is also used to describe synthetic or fusion molecules encoding all or part of an expression product. In particular embodiments, the term "nucleic acid molecule" and "gene" may be used interchangeably.

In a particularly useful embodiment, the present invention provides a promoter for the *WARP* gene. The identification of the promoter permits ECM and in particular cartilage-specific expression of particular genetic sequences. The latter would include a range of therapeutic molecules such as cytokines, growth factors, antibiotics or other molecules to assist in the treatment of disease, trauma or other conditions of the ECM.

Accordingly, another aspect of the present invention provides a cartilage-specific promoter or functional derivative or homologue thereof, said promoter *in situ* operably linked to a nucleotide sequence comprising SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms or a nucleotide sequence having at least about 65% similarity to SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or SEQ ID NO:5 or their complementary forms under low stringency conditions.

The promoter is conveniently resident in a vector which comprises unique restriction sites to facilitate the introduction of genetic sequences operably linked to said promoter.

The present invention further contemplates a genetically modified animal.

More particularly, the present invention provides an animal model useful for screening for agents capable of ameliorating the effects of compromised ECM and in particular cartilage. In one embodiment, the animal model produce low amounts of WARP. Such an animal

would havc a predisposition for ECM-mediated diseases. Such an animal model is useful for screening for agents which ameliorate such conditions.

Accordingly, another aspect of the present invention provides a genetically modified animal wherein said animal produces low amounts of WARP relative to a non-genetically modified animal of the same species.

Preferably, the genetically modified animal is a mouse, rat, guinea pig, rabbit, pig, sheep or goat. More preferably, the genetically modified animal is a mouse or rat. Most preferably, the genetically modified animal is a mouse.

Accordingly, a preferred aspect of the present invention provides a genetically modified mouse wherein said mouse produces low amounts of WARP relative to a non-genetically modified mouse of the same strain.

The animal model contemplated by the present invention comprises, therefore, an animal which is substantially incapable of producing a WARP. Generally, but not exclusively, such an animal is referred to as a homozygous or heterozygous WARP-knockout animal. Such animals exhibit ECM-mediated disease conditions. These animals are useful for screening for agents which ameliorate such conditions and which can reduce the clinical severity of the disease condition. Once such molecules are identified, a treatment or prophylactic protocol can be developed which targets these conditions.

The animal models of the present invention may be in the form of the animals or may be, for example, in the form of embryos for transplantation. The embryos are preferably maintained in a frozen state and may optionally be sold with instructions for use.

Yet another aspect of the present invention provides a targeting vector useful for inactivating a gene encod *WARP* said targeting vector comprising two segments of genetic material encoding said *WARP* flanking a positive selectable marker wherein when said targeting vector is transfected into embryonic stem (ES) cells and the marker selected, an

ES cell is generated in which the gene encoding said *WARP* is inactivated by homologous recombination.

Still another aspect of the present invention provides a targeting vector useful for inactivating a gene encoding WARP, said targeting vector comprising two segments of genetic material encoding WARP flanking a positive selectable marker wherein when said targeting vector is transfected into embryonic stem (ES) cells and the marker selected, an ES cell is generated in which the *WARP* gene is inactivated by homologous recombination.

Preferably, the ES cells are from mice, rats, guinea pigs, pigs, sheep or goats. Most preferably, the ES cells are from mice.

Still yet another aspect of the present invention is directed to the use of a targeting vector as defined above in the manufacture of a genetically modified animal substantially incapable of producing WARP.

Even still another aspect of the present invention is directed to the use of a targeting vector as defined above in the manufacture of a genetically modified mouse substantially incapable of producing WARP.

Preferably, the vector is DNA. A selectable marker in the targeting vector allows for selection of targeted cells that have stably incorporated the targeting DNA. This is especially useful when employing relatively low efficiency transformation techniques such as electroporation, calcium phosphate precipitation and liposome fusion where typically fewer than 1 in 1000 cells will have stably incorporated the exogenous DNA. Using high efficiency methods, such as microinjection into nuclei, typically from 5-25% of the cells will have incorporated the targeting DNA; and it is, therefore, feasible to screen the targeted cells directly without the necessity of first selecting for stable integration of a selectable marker. Either isogenic or non-isogenic DNA may be employed.

Examples of selectable markers include genes conferring resistance to compounds such as antibiotics, genes conferring the ability to grow on selected substrates, genes encoding proteins that produce detectable signals such as luminescence. A wide variety of such markers are known and available, including, for example, antibiotic resistance genes such as the neomycin resistance gene (*neo*) and the hygromycin resistance gene (*hyg*). Selectable markers also include genes conferring the ability to grow on certain media substrates such as the *tk* gene (thymidine kinase) or the *hprt* gene (hypoxanthine phosphoribosyltransferase) which confer the ability to grow on HAT medium (hypoxanthine, aminopterin and thymidine); and the bacterial *gpt* gene (guanine/xanthine phosphoribosyltransferase) which allows growth on MAX medium (mycophenolic acid, adenine and xanthine). Other selectable markers for use in mammalian cells and plasmids carrying a variety of selectable markers are described in Sambrook *et al.*, 1989 [57].

The preferred location of the marker gene in the targeting construct will depend on the aim of the gene targeting. For example, if the aim is to disrupt target gene expression, then the selectable marker can be cloned into targeting DNA corresponding to coding sequence in the target DNA. Alternatively, if the aim is to express an altered product from the target gene, such as a protein with an amino acid substitution, then the coding sequence can be modified to code for the substitution, and the selectable marker can be placed outside of the coding region, for example, in a nearby intron.

The selectable marker may depend on its own promoter for expression and the marker gene may be derived from a very different organism than the organism being targeted (e.g. prokaryotic marker genes used in targeting mammalian cells). However, it is preferable to replace the original promoter with transcriptional machinery known to function in the recipient cells. A large number of transcriptional initiation regions are available for such purposes including, for example, metallothionein promoters, thymidine kinase promoters, β-actin promoters, immunoglobulin promoters, SV40 promoters and human cytomegalovirus promoters. A widely used example is the pSV2-*neo* plasmid which has tahe bacterial neomycin phosphotransferase gene under control of the SV40 early promoter and confers in mammalian cells resistance to G418 (an antibiotic related to neomycin). A

number of other variations may be employed to enhance expression of the selectable markers in animal cells, such as the addition of a poly(A) sequence and the addition of synthetic translation initiation sequences. Both constitutive and inducible promoters may be used.

The DNA is preferably modified by homologous recombination. The target DNA can be in any organelle of the animal cell including the nucleus and mitochondria and can be an intact gene, an exon or intron, a regulatory sequence or any region between genes.

Homologous DNA is a DNA sequence that is at least 70% identical with a reference DNA sequence. An indication that two sequences are homologous is that they will hybridize with each other under stringent conditions [57].

The present invention is further described by the following non-limiting Examples.

EXAMPLE 1

Identification of WARP cDNAs

The mouse EST database was conceptually translated into six reading frames and interrogated with the 200 amino acid sequence of the α3(VI) N8 VA domain of human collagen VI [3] using the tblastn program (v2.0) at the National Center for Biotechnology Information (NCBI). Several overlapping cDNA clones with significant similarity to α3(VI) N8 at the protein level were identified. The inventors obtained three of these clones, ui42d08, ue22e08 and ml15f02 from E12.5 mouse embryo, spleen and kidney, respectively (Genome Systems). DNA sequencing (Amplicycle sequencing kit, Perkin Elmer Biosystems) revealed that clones ue22e08 (1026 bp) and mt15f02 (551 bp) lie entirely within the ui42d08 (2308 bp) sequence and exactly matched the larger clone spanning nucleotides 1282-2308 and 1833-2227, confirming that the three cDNAs represent a single gene.

EXAMPLE 2

WARP plasmid constructs and expression in transfected cells

The ui42d08 cDNA in pME18 (GenBank Accession number AI115125) (Figure 1A) was subcloned into the pBluescriptSK$^-$ vector (Stratagene) as a *Xho*1 fragment. The clone was then sequenced using the Amplicycle sequencing kit (Perkin Elmer Biosystems) and translated *in vitro* using the TNT Coupled Transcription and Translation System (Promega) [28] to confirm the open reading frame. To generate a WARP GST-VA domain fusion construct, the VA domain sequence from amino acid 21-212 was amplified by PCR using primers that anneal in the cDNA sequence between nucleotides 92-111 and 648-666. The primers were designed to include flanking *Bam*H1 and *Eco*R1 sites to allow in-frame cloning of the VA domain PCR product into the glutathione S-transferase fusion vector pGEX-2T (Amersham Pharmacia). To enable immunoprecipitation of WARP protein from transfected cells, a His-tagged full-length expression construct was also produced. Six histidine residues were incorporated at the N-terminus immediately following amino acid 21, between the signal peptide and the start of the VA domain, by strand overlap extension

PCR [28] and subcloned into the pBluescriptSK$^-$ vector. To allow episomal expression in mammalian cells, *WARP*-His was subcloned from pBluescriptSK$^-$ into pCEP4 (InVitrogen) as a *Xho*1 fragment. *WARP*-His in pCEP4 was transfected into 293-EBNA cells (InVitrogen) grown in Dulbecco's Modified Eagles Medium (DMEM) containing 10% v/v bovine serum using FuGene transfection reagent (Boehringer Mannheim) according to the manufacturer's instructions and grown for 14 days in the presence of 250 μg/ml hygromycin B (Boehringer Mannheim) to select for transfected cells.

EXAMPLE 3

Cell culture

Human embryonic kidney 293-EBNA cells, mouse MC3T3 osteoblast [29], Mov13 fibroblast [30] and C2C12 myoblast [31] cell lines were maintained in culture in DMEM containing 10% v/v bovine serum. Primary chondrocytes were isolated as previously described [32]. Briefly, rib cages were dissected from newborn mice and incubated in DMEM containing 5% v/v bovine serum and 2 mg/ml collagenase (Worthington Biochemical Corp.) for 30 mins at 37°C. Loose connective tissue and bone was removed and the rib cartilage incubated in fresh collagenase solution for 16 hrs. Chondrocytes released from cartilage were either centrifuged to pellet cells or plated out as a monolayer in a 60-mm dish. Pelleted cells, which retained a chondrocyte phenotype, were grown in DMEM containing 10% w/v fetal calf serum for 16 hrs prior to RNA isolation. Cells grown as a monolayer were cultured for 48 hrs prior to RNA isolation to allow chondrocyte de-differentiation [32]. Mouse MCT chondrocytes, immortalized with a temperature sensitive SV-40 large T-antigen [33], were cultured at the permissive temperature of 32°C, where the cells proliferate and express an osteoblast-like phenotype as demonstrated by expression of the osteoblast markers type I collagen and bone Gla protein. When grown at the non-permissive temperature of 37°C, the cells cease dividing and express type X collagen, matrix Gla protein and osteopontin, which are markers of hypertrophic chondrocytes. For one experiment MCT cells were grown at 37°C for 3 days to induce a hypertrophic-like phenotype then transferred to 32°C for 3 days to induce an osteoblast-like phenotype.

EXAMPLE 4

mRNA expression analysis

Total RNA was isolated from mouse cell lines and primary rib chondrocytes using the mini Rneasy (trademark) RNA isolation kit (Qiagen) according to the manufacturer's instructions and from mouse tissues using the guanidinium thiocyanate and phenol/chloroform method of Chomzynski [34]. To ensure that no genomic DNA was carried through the isolation procedure all RNA samples were digested with DNA-free (trademark) DNase Treatment and Removal kit (Ambion) and repurified using the Rneasy (trademark) kit. Each sample was then assessed for genomic DNA contamination by performing a RT-PCR reaction in the absence of reverse transcriptase. *WARP mRNA* expression was determined by Northern blot analysis, RT-PCR and semi-quantitative RT-PCR. For Northern blot analysis, 60 μg of total RNA was poly(A)-selected using oligo dT Dynabeads (Dynal), fractionated on a 1% w/v agarose formaldehyde gel and transferred to Hybond N^+ nylon membrane (Amersham). A [^{32}P]-dCTP-labeled *WARP* probe was hybridized to the blot in Ultrahyb hybridization solution (Ambion) at 42°C overnight. The blot was washed to a stringency of 0.1 x SSC/0.1% w/v SDS at 65°C and subjected to autoradiography. RT-PCR was performed using the GeneAmpR RNA PCR kit (Perkin Elmer). Two μg of total RNA was added to each RT reaction in a total volume of 40 μl and 10 μl of cDNA was used in the subsequent PCR in a 50 μl reaction volume. The optimal Mg^{2+} concentration was found to be 0.35 mM for the *WARP* amplification and 1 mM for the internal control, hypoxanthine guanine phosphoribosyltransferase (HPRT), a housekeeping gene involved in purine metabolism. In the PCR step, NR1 [^{1666}5'-CTCAAAGCCATGCGTAGTCC-3'^{1685} (SEQ ID NO:9)], and NF4 [^{953}5'-AGAACGCATCGTCATCTCGC-3'^{972} (SEQ ID NO:10)] primers were used to amplify a 693 bp region of *WARP*. mHPRT1 [^{231}5'-CCTGCTGGATTACATTAAAG-3'^{251} (SEQ ID NO:11)] and mHPRT2 [^{581}5'-TCAAGGGCATATCCAACAAC-3'^{601} (SEQ ID NO:12)] primers were used to amplify a 350 bp fragment of the mouse HPRT gene (GenBank Accession Number NM_013556). The cycle number for each gene was selected so that amplification was in the linear range, allowing the level of PCR products to be compared between samples. Simultaneous amplification of HPRT derived from the same cDNA

reaction allowed correction for small variations in amount of template. For RT-PCR, primers and probes were designed with Primer Express (v1.0) software according to Applied Biosystems guidelines, and obtained directly from Applied Biosystems. The fluorophores, carboxyfluorescein (FAM) and VIC (trademark) were added to the 5' end of *WARP* and HPRT probes respectively, and the N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA) fluorophore added to the 3' end of both probes during synthesis. The *WARP* probe [5'-(FAM)-CTGGTCATCGCCGCCCTTGC-(TAMRA)-3' (SEQ ID NO:13)] and primers [^{1399}5'-GACCAGCGTTAATTCCTTTCGT-3' (SEQ ID NO:14) and 5'-CCGGGTTTCCCGGAAGT-3'^{1472} (SEQ ID NO:15) amplified a 73 bp region. The HPRT probe [5'-(VIC)-TTACTGGCAACATCAACAGGACTCCTCGTATT-(TAMRA)-3' (SEQ ID NO:16)] and primers [^{739}5'-CCACAGGACTAGAACACCTGCTAA-3' (SEQ ID NO:17) and 5'-CCTAAGATGAGCGCAAGTTGAA-3'^{825} (SEQ ID NO:18) amplified an 86 bp region. In the intact probe, TAMRA is able to quench FAM and VIC but during the PCR the reporter fluorophores are released into solution by the 5'-exonuclease activity of the polymerase allowing them to fluoresce. The amount of fluorescence is directly proportional to the amount of specific product generated in the PCR. Reactions were performed on a Perkin Elmer Life Sciences ABI PRISM 7700 Sequence Detector using the TaqMan Universal PCR master mix (Applied Biosystems) containing AmpliTaq Gold polymerase and repeated several times with similar results. The data are expressed as a ratio of *WARP*:HPRT mRNA at a cycle number that falls within the linear range of amplification as determined by visual examination of the data generated by Sequence Detector (v1.7) software (Applied Biosystems).

EXAMPLE 5

Production of an anti-WARP antibody

The GST-VA fusion cDNA construct in pGEX-2T was transformed into competent DH5α bacteria, individual colonies grown and fusion protein expression induced by IPTG [35]. The insoluble fusion protein was purified from cell preparations using a Mini Whole Gel Eluter Harvester (BioRad) and injected into a NZ White rabbit. Antisera from the rabbit

immunised with purified GST-VA domain fusion protein bound to the fusion protein in a dose dependent manner in an ELISA assay. To demonstrate specificity of the antibody for WARP, the fusion protein was cleaved with thrombin to separate the GST and VA domains and subjected to immunoblotting using the antisera as probe. The antisera recognised both GST and the VA domain at a dilution of 1 in 1000.

EXAMPLE 6

Cartilage sample preparation and Western blotting

Joint and rib tissue was dissected from newborn mice and cleaned of surrounding bone and connective tissue. Cartilage samples were powdered in a freezer mill (Spex) and dissolved in extraction solution 1 (40 mM Tris/HCl, pH 7.5, 10 mM EDTA containing 'Complete' protease inhibitors (Roche)). Samples were then vortexed and sonicated for 20 secs and the insoluble material pelleted in a microcentrifuge. The supernatant was collected and saved as soluble fraction 1 and the insoluble pellet washed and sonicated three times in Tris/HCl, pH 7.5, 10 mM EDTA. The pellet was resuspended in extraction solution 1 and treated overnight at 37°C with 0.02 units of chondroitinase ABC (ICN) and 1 unit of hyaluronidase (Sigma). Samples were pelleted and washed three times with 40 mM Tris/HCl, pH 7.5, 10 mM EDTA and the supernatants saved as soluble fraction 2. The remaining pellet was dissolved in 6 M GuHCl, 40 mM Tris/HCl, pH 7.5, 10 mM EDTA containing protease inhibitors for 5 hrs at 4°C, then centrifuged. The supernatant was saved as soluble fraction 3 and the matrix components precipitated with 95% v/v ethanol and the pellet washed with 70% v/v ethanol. Samples were then freeze-dried and resuspended in 200 μl of 8 M urea, 4% v/v cholamidopropyl-dimethylammonio-propane-sulfonate (CHAPS), 40 mM Tris-HCl, pH 7.5, containing 2 mM tributylphosphine and 2.5% v/v β-mercapto-ethanol. For some experiments the reducing agents were omitted.

The protein content of extracts 1, 2, and 3 was determined by the Bradford assay and 20 μg total protein aliquots were denatured by heating at 95°C for 5 min, separated on a 10% w/v SDS-polyacrylamide gel and transferred to Immobilon (trademark)-P PVDF membrane (Millipore). The membrane was blocked in 5% w/v milk powder in PBS for 1

hr and then incubated in antibody buffer (0.5% w/v milk powder in 0.1% w/v Tween-20 in PBS) containing either WARP or matrilin-1 antisera [36] (1 in 1000 and 500 dilution, respectively) for 1 hr at room temperature. Following three washes in 0.1% w/v Tween-20 in PBS, anti-rabbit IgG-HRP secondary antibody (Dako Corporation) was added at a dilution of 1 in 10,000 in antibody buffer and incubated for 1 hr. Following washing, the signal was developed with ECL Plus Western blotting detection system (Amersham Pharmacia) and autoradiography performed using X-OMAT film (Kodak).

EXAMPLE 7

WARP biosynthetic labeling and analysis

293-EBNA cells transfected with *WARP*-His cDNAwere grown to confluence in a 60-mm dish and labeled for 16 hrs with 300 μCi of L-[^{35}S]-methionine (1398 Ci/mmol, NEN Research Products) in DMEM without L-methionine and L-cysteine (Life Technologies, Inc) as previously described [26]. The medium fraction was removed and clarified centrifuged and NP-40 added to the supernatant to 1% v/v together with a cocktail of protease inhibitors (1 mM 4-(2 aminoethyl)-benzenesulfonyl-flouride (AEBSF); 1 mM phenylmethylsulfonyl fluoride (PMSF); 20 mM N-ethylmaleimide (NEM)). The cell layer was dispersed in 1ml of lysis buffer (150 mM NaCl; 50 mM Tris-HCl, pH 7.5; 5 mM EDTA; 20 mM NEM; 1 mM AEBSF; 1 mM PMSF; 1% v/v NP-40) on ice for 30 min. then centrifuged briefly to remove insoluble material. Following a pre-clear step with 100 μl protein G-sepharose (20% w/v slurry in PBS), anti-His antibody (Boehringer Mannheim)(1 in 100 dilution) was added to each fraction together with 100 μl fresh protein G-sepharose and mixed gently at 4°C for 16 hrs. The antibody-sepharose complex was washed twice with 50% w/v lysis buffer/50% w/v NET (150 mM NaCl; 50 mM Tris-HCl, pH 7.4; 1 mM EDTA; 0.1% w/v NP-40) for 30 min each then twice with NET. Immunoprecipitated material was separated from the sepharose beads by heating at 65°C for 15 min in SDS-PAGE sample buffer containing 20 mM dithiothreitol (DTT), fractionated on a 10% w/v) SDS-polyacrylamide gel and subjected to fluorography.

EXAMPLE 8

N-glycosidase treatment

WARP-His protein was deglycosylated by N-glycosidase F (Roche) treatment according to the manufacturer's guidelines. Immunoprecipitated WARP-His was denatured by boiling in 1% w/v SDS for 2 min then diluted 1 in 10 with sodium phosphate buffer (20 mM sodium phosphate, pH 7.2; 10 mM sodium azide; 50 mM EDTA; 0.5% v/v NP-40) and boiled again for 2 min Following addition of 0.4 units of N-Glycosidase F the sample was incubated for 20 hrs at 37°C then heat denatured in sample buffer containing 20 mM DTT and analysed by SDS-polyacrylamide gel electrophoresis.

EXAMPLE 9

SDS-polyacrylamide gel electrophoresis

Samples were resolved on 10% w/v polyacrylamide separating gels with a 3.5% w/v acrylamide stacking gel in the absence of urea as described previously [37]. Prior to electrophoresis, samples were diluted with loading buffer to give a final concentration of 0.125 mM Tris/HCl, pH 6.8 containing 2% w/v SDS and denatured for 10 min or otherwise indicated. Electrophoresis conditions and fluorography of radioactive gels have been described previously [28,37].

EXAMPLE 10

Immunohistochemistry

Newborn mouse hind limbs were surgically removed and frozen in OCT compound (Sakura). 10 μM sections were cut and fixed in 95% v/v methanol/5% v/v acetic acid on ice for 10 min. To facilitate antibody penetration into the ECM, sections were treated with 0.2% w/v hyaluronidase in PBS for 20 min at room temperature. Following a 5 min wash in PBS the sections were treated with 0.3% H_2O_2/0.3% w/v serum in PBS for 5 min to inactivate endogenous peroxidases. The sections were stained with the WARP antibody (1

in 100 dilution) using the Vectastain Elite ABC kit (Vector Laboratories) and colour was developed using a DAB peroxidase substrate kit for Vectastain.

EXAMPLE 11

WARP

To identify novel ECM proteins that contain VA-like domains, the mouse EST database at the NCBI was searched with the N-terminal N8 VA domain of the $\alpha3$ chain of human collagen VI [3]. The inventors identified several overlapping EST clones that when fully sequenced clearly represent a novel gene that contains a predicted VA-like protein module. The longest EST clone, ui42d08, appeared to be a full-length cDNA with a start methionine codon at nucleotides 30-32 and an in-frame stop TGA codon at 1275-1277, indicating an open reading frame of 1248 bps with 29 bps of 5'UTR and 1063 bps of 3'UTR (Figure 1A). The 3' end of the clone includes a poly(A) tail and a predicted polyadenylation site at nucleotides 2279-2285. The full-length WARP cDNA was transcribed and translated *in vitro* and SDS-PAGE analysis demonstrated a single protein product with an apparent molecular weight of 55 kDa indicating that no stop codons were present within the open reading frame. Since the full-length *WARP* nucleotide and protein sequences have not been previously reported and the VA domain is related to, but distinctly different from, those described in existing family members, the inventors conclude that this gene is a new member of the VA superfamily. The inventors named this gene, WARP, for von Willebrand factor A-domain related protein.

The *WARP* open reading frame encodes a 415 amino acid protein with a predicted molecular weight of 45 kDa. An 18 amino acid signal sequence with a cleavage site between Ala^{18} and Arg^{19} is indicated by signal sequence prediction program SignalP (v2.0) (http://genome.cbs.dtu.dk/services/SignalP-2.0) [38]. The signal sequence is followed by a VA domain of approximately 200 amino acids with a putative metal ion-dependent adhesion site (MIDAS) [12] and three potential O-linked sites at Ser^{148}, Thr^{362} and Thr^{401}, as predicted by NetOGlyc software (http://genome.cbs.dtu.dk/services/NetOGlyc) [39](Figure 1A). Adjacent to the VA domain are two fibronectin type III (F3) repeats of

approximately 80 amino acids in length each containing a potential N-linked glycosylation site at Asn^{264} and Asn^{359} that fits the consensus sequence NxS/T. The 21 amino acid C-terminus at the end of the second F3 repeat is rich in proline and arginine residues but did not show homology to any other protein by extensive database searching (BLASTP v2.1.2). The domain structure of WARP is shown in Figure 1B.

EXAMPLE 12

Similarity of WARP to other ECM proteins

The protein sequences of the two protein domains present in WARP (VA and F3) were used to search the Non-Redundant and Conserved Domain databases at NCBI. A high degree of amino acid similarity exists between the WARP VA domain and those found in other proteins with most similarity to VA domains present in the FACIT collagens XII, XIV (Figure 2A) and the recently described FACIT collagen XX. The amino acids within the MIDAS motif which are critical for ion binding, Asp^{40}, Ser^{42}, Ser^{44}, Thr^{113} and Asp^{144} are conserved in WARP although biochemical and crystallographic studies are required to directly demonstrate a functional MIDAS motif. In addition, the overall arrangement of alpha helices and beta sheets that form the important secondary structural framework that is shared between all VA-like domains is conserved in WARP. The two F3 repeats are less conserved than the VA domain, although the overall framework of 7 hydrophobic strands that form the β-sandwich typical of F3 repeats is conserved [40](Figure 2B). The first F3 repeat, F3-1, is most similar to those found in tenascins and collagen XIV and F3-2 is most similar to those in collagen VII and the FACIT collagens.

To determine whether the predicted signal sequence is functional in directing WARP secretion from cells, and to determine if the putative N-glycosylation sites are utilized, a *WARP* cDNA expression construct with a poly-His tag inserted between the signal peptide and VA domain was transfected into 293-EBNA fibroblasts. The stably transfected cells were labeled overnight with ^{35}S-methionine and immunoprecipitated with anti-His antibodies. No material was immunoprecipitated from untransfected 293-EBNA cells (Figure 3, lanes 1 and 2) indicating that no endogenous proteins are recognised by the anti-

His antibody. In cells transfected with the *WARP*/His cDNA, His-tagged WARP protein was present as an approximately 55 kDa band in cell layer fractions both the media and band (lanes 3 and 4). The majority of WARP is detected in the medium during these continuous labeling conditions, suggesting that WARP is efficiently secreted from cells and functions in the ECM environment. When the immunoprecipitated protein is subjected to N-glycosidase digestion there is a mobility shift to approximately 53 kDa indicating that WARP has one or more N-linked sugar side chains (lane 5). There are two possible N-glycosylation sites at Asn^{254} and Asn^{359} located in similar positions in the centre of each of the two F3 repeats in a loop region between β-strands C and C' (Figure 2B).

EXAMPLE 13

WARP mRNA is expressed highest in chondrocytes

The *WARP mRNA* expression pattern in cell lines was examined by Northern blot analysis using poly(A) mRNA selected from primary rib chondrocytes, Mov13 fibroblasts, MC3T3 osteoblasts and C2C12 myoblasts (Figure 4A). *WARP mRNA* was present in chondrocytes (lane 1) but not in the osteoblast, fibroblast and myoblast cell lines (lanes 2-4). *WARP* migrates as a 2.3 kb mRNA which is in agreement with the size of the full-length *WARP* cDNA represented by clone ui42d08 which is 2308 bp in size (see Figure 1).

To examine the expression of *WARP mRNA* in a wider range of tissues, total RNA was isolated from mouse heart, skeletal muscle, testis, brain, and lung, and subjected to RT-PCR using primers specific for *WARP* and a control, HPRT (Figure 4B). To control for variation between RT reactions, *WARP* and HPRT were amplified in separate reactions using the same template cDNA. Following 36 cycles of amplification, a *WARP* PCR product was present in chondrocyte RNA (upper panel, lane 6) but not in any other tissues or cell lines. The presence of a band representing HPRT in all lanes (lower panel) indicates that for all samples the starting RNA was intact and the RT reactions were successful.

To gain a reliable and semi-quantitative estimation of *WARP mRNA* levels in chondrocytes and cell lines a third technique for assaying mRNA levels, Real-time PCR, was employed

(Figure 4C). In this method, a fluorescently-labeled probe, designed to anneal between two opposing primers, is removed by the action of the polymerase allowing an accurate estimation of PCR product levels by the appearance of a fluorescent signal in solution. By labeling each probe with a different fluorophore, the amplification reaction can be performed in the same tube, which controls for variations in amount of input cDNA and in the efficiency of the amplification reaction. The data are expressed as a ratio of *WARP*:HPRT mRNA at a cycle number that falls within the linear range of amplification. *WARP mRNA* levels were 7-fold higher in both primary rib chondrocytes and MCT cells induced to form a hypertrophic chondrocyte-like phenotype compared to MCT cells induced to form an osteoblast-like phenotype and MC3T3 osteoblasts. Expression in chondrocytes was >20-fold higher compared to fibroblasts cell lines and fibroblast-like cells derived from de-differentiated primary chondrocytes. These differences in the level of *WARP* expression are consistent with those detected by Northern analysis (Figure 4A) and RT-PCR (Figure 4B) and indicate that *WARP* is expressed highest in chondrocytes and at much lower levels in other tissues and cell lines.

These expression experiments demonstrate that *WARP mRNA* is expressed highest in primary rib chondrocytes which contain a mixed population of resting, proliferative, maturing and hypertrophic chondrocytes and in MCT cells induced to express a hypertrophic chondrocyte-like phenotype [33]. *WARP mRNA* was undetected or expressed at very low levels in all other tissues and cell lines examined including MCT cells induced to form osteoblast-like cells. Interestingly, *WARP* expression was down-regulated when rib chondrocytes were allowed to de-differentiate into fibroblast-like cells suggesting that expression is tightly controlled by the chondrocyte program of gene expression. This is supported by our finding that when MCT cells are induced to change from a hypertrophic-like to an osteoblast-like phenotype by changing the temperature of incubation from 37°C to 32°C, *WARP* expression was reduced approximately 6-fold (Figure 4C).

EXAMPLE 14

WARP protein expression in cartilage

To detect WARP protein *in vivo,* a polyclonal antibody against a bacterially expressed GST-VA domain fusion protein was made and used to probe an immunoblot containing serial extractions of newborn cartilage. When cartilage was extracted with Tris-buffered EDTA, either before (F1 extract) or after degradation of the aggrecan complex and the glycosaminoglycan side chains with chondroitinase and hyaluronidase (F2 extract), and resolved by SDS-PAGE under reducing conditions, no WARP protein was detected (Figure 5A, lanes 2 and 3). These data suggest that that WARP was neither a soluble matrix component nor one that interacts with the matrix *via* divalent cation-dependant mechanisms. When cartilage was further extracted under denaturing conditions with 6 M guanidine (F3 extract), a strong WARP band was detected (Figure 5A, lane 4). Under these extraction conditions matrilin-1 was also present exclusively in the guanidine extract (Figure 5A. lane 5). Previous data have shown that matrilin-1 occurs in several pools of increasing insolubility in cartilage [41]. One pool is released by buffered EDTA containing 0.25 M NaCl, a second pool, which is strongly associated with aggrecan, requires chaotrophic agents for dissociation, and a third pool that increases with cartilage maturation is covalently linked to aggrecan, part of which can be released by reduction under denaturing conditions.

The inventors clearly show that WARP is also found in the cartilage matrix *in vivo*, and the necessity for extraction with a chaotrophic agent suggests that it may be a strongly interacting matrix component. However, the experiments do not provide insight on whether WARP also exists in a number of pools of differing solubilities and possibly different functions during development or maturation. A proportion of WARP may also be present as insoluble supramolecular aggregates or covalently linked to guanidine-insoluble matrix components. These important questions will be addressed by further detailed biochemical analysis.

To determine the location of WARP protein in cartilage, sagittal sections of newborn mouse tibia were subjected to immunohistochemistry using the WARP antibody (Figure 5B). WARP stained the extracellular space surrounding chondrocytes in all zones of cartilage of the tibial head. The signal was relatively uniform throughout the zones although staining was more intense in the hypertrophic zone compared to the neighbouring pre-hypertrophic zone (right panel). In general, the signal was strongest in the pericellular regions with the matrix furthest from chondrocytes showing relatively less staining for WARP. In control samples where the WARP antibody was omitted, no staining was present. The inventors conclude from these experiments that WARP is expressed by chondrocytes and is a component of the ECM surrounding chondrocytes in the cartilage of newborn mice. The presence of WARP protein throughout all zones of developing cartilage is similar to that of other structural matrix proteins including matrilin-1, matrilin-3, aggrecan, collagen II and COMP [42-44] suggesting that WARP is a fundamental component of the cartilage ECM.

EXAMPLE 15

WARP is an oligomer in vivo

To determine whether WARP exists as a monomer or forms higher-order structures *in vivo*, guanidine-soluble extracts were prepared from newborn mouse rib cartilage and subjected to SDS-PAGE analysis under reducing and non-reducing conditions and immunoblotted using WARP antisera (Figure 6). When cartilage extracts were prepared and resolved under reducing conditions WARP migrated as a 55 kDa monomer (Figure 5A, lane 4) although in some experiments there was also some higher-order oligomeric forms of WARP (Figure 6, lane 1). These are presumably due to incomplete reduction or dissociation during sample preparation. In contrast, when the cartilage extract was prepared and fractionated in the absence of reducing agents WARP was present exclusively as higher-order oligomers and there was a complete absence of 55 kDa monomeric WARP (lane 2). The WARP oligomer migrates as a smeared band (Figure 6, lane 2), which may reflect variability in the numbers of WARP monomers in the oligomer, or possibly variation in the glycosylation pattern of WARP monomers which also

demonstrate a diffuse electrophoretic migration (Figure 5A, lane 4 and Figure 6, lane 2). These experiments clearly demonstrate that endogenous WARP forms disulfide-bonded multimers of greater than 200 kDa in size although it is not known whether these are composed of WARP homo-oligomers, or hetero-oligomers where WARP is disulfide bonded to other ECM proteins.

The C-terminus of matrilin-1 forms a coiled-coil structure composed of a heptad repeat of hydrophobic amino acids which directs the formation of matrilin multimers [46]. Multimers are then stabilized by interchain disulfide bonds provided by two Cys residues present within the C-terminus [47]. The C-terminal domain in WARP is not predicted to form a coiled-coil structure of the type found in matrilins because it does not contain a well defined heptad repeat of hydrophobic residues. However, the C-terminal Cys residues, at Cys^{369} and Cys^{393} in the second F3 repeat, would be in a good position to stabilize WARP oligomerisation and it is tempting to speculate that the C-terminus of WARP is involved in the formation of WARP oligomers.

EXAMPLE 16

Human WARP

A human homologue of murine *WARP* was identified by database homology searching. The nucleotide sequence (SEQ ID NO:5) and corresponding amino acid sequence (SEQ ID NO:6) are shown in Figure 6.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

BIBLIOGRAPHY

1. Engel, J., Efimov, V. P., and Maurer, P. (1994) *Development* **Suppl.,** 35-42

2. Colombatti, A., Bonaldo, P., and Doliana, R. (1993) *Matrix* **13,** 297-306

3. Chu, M. L., Zhang, R. Z., Pan, T. C., Stokes, D., Conway, D., Kuo, H. J., Glanville, R., Mayer, U., Mann, K., Deutzmann, R., and Timpl, R. (1990) *EMBO J.* **9,** 385-393

4. Chu, M. L., Pan, T. C., Conway, D., Kuo, H. J., Glanville, R. W., Timpl, R., Mann, K., and Deutzmann, R. (1989) *EMBO J.* **8,** 1939-194

5. Parente, M. G., Chung, L. C., Ryynanen, M., Woodley, D. T., Wynn, K. C., Bauer, E. A., Mattei, M.-G., Chu, M.-L., and Uitto, J. (1991) *Proc.Natl.Acad.Sci.U.S.A.* **88,** 6931-6935

6. Yamagata, M., Yamada, K. M., Yamada, S. S., Shinomura, T., Tanaka, H., Nishida, Y., Obara, M., and Kimata, K. (1991) *J.Cell Biol.* **115,** 209-221

7. Trueb, J. and Trueb, B. (1992) *Eur.J.Biochem.* **207,** 557

8. Koch, M., Foley, J. E., Hahn, R., Zhou, P., Burgeson, R. E., Gerecke, D. R., and Gordon, M. K. Papers in Press: J. Biol. Chem. published March 23, 2001 as 10.1074/jbc.M009912200

9. Deak, F., Wagener, R., Kiss, I., and Paulsson, M. (1999) *Matrix Biol.* **18,** 55-64

10. Robertson, N. G., Skvorak, A. B., Yin, Y., Weremowicz, S., Johnson, K. R., Kovatch, K. A., Battey, J. F., Bieber, F. R., and Morton, C. C. (1997) *Genomics* **46,** 345-354

11. Gilges, D., Vinit, M. A., Callebaut, I., Coulombel, L., Cacheux, V., Romeo, P., and Vigon, I. (2000) *Biochem.J.* **352,** 49-59

12. Lee, J.-O., Rieu, P., Arnaout, M. A., and Liddington, R. (1995) *Cell* **80,** 631-638

13. Mole, J. E. (1984) *J.Biol.Chem.* **259,** 3407-3412

14. Chan, P., Risler, J. L., Raguenez, G., and Salier, J. P. (1995) *Biochem.J.* **306,** 505-512

15. Ellis, S. B., Williams, M. E., Ways, N. R., Brenner, R., Sharp, A. H., Leung, A. T., Campbell, K. P., McKenna, E., Koch, W. J., Hui, A., Schwartz, A., and Harpold, M. M. (1988) *Science* **241,** 1661-1664

16. Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., Harlan, J. M., Titani, K., and Davie, E. W. (1985) *Proc.Natl.Acad.Sci.U.S.A.* **82,** 6394-6398

17. Emsley, J., Cruz, M., Handin, R., and Liddington, R. (1998) *J.Biol.Chem.* **273,** 10396-10401

18. Bienkowska, J., Cruz, M., Atiemo, A., Handin, R., and Liddington, R. (1997) *J.Biol.Chem.* **272,** 25162-25167

19. Qu, A. and Leahy, D. J. (1995) *Proc.Natl.Acad.Sci.U.S.A.* **92,** 10277-10281

20. Emsley, J., King, S. L., Bergelson, J. M., and Liddington, R. C. (1997) *J.Biol.Chem.* **272,** 28512-28517

21. Ruggeri, Z. M. (1997) *J.Clin.Invest.* **99,** 559-564

22. Denis, C., Baruch, D., Kielty, C. M., Ajzenberg, N., Christophe, O., and Meyer, D. (1993) *Arteriosclerosis & Thrombosis* **13,** 398-406

23. Tuckwell, D. S., Reid, K. B., Barnes, M. J., and Humphries, M. J. (1996) *Eur.J.Biochem.* **241,** 732-739

24. Specks, U., Mayer, U., Nischt, R., Spissinger, T., Mann, K., Timpl, R., Engel, J., and Chu, M.-L. (1992) *EMBO J.* **11,** 4281-4290

25. Kuo, H.-J., Maslen, C. L., Keene, D. R., and Glanville, R. W. (1997) *J.Biol.Chem.* **272,** 26522-26529

26. Fitzgerald, J., Morgelin, M., Selan, C., Wiberg, C., Keene, D. R., Lamande, S. R., and Bateman, J. F. (2001) *J.Biol.Chem.* **276,** 187-193

27. Chen, Q., Zhang, Y., Johnson, D. M., and Goetinck, P. F. (1999) *Mol.Biol.Cell* **10,** 2149-2162

28. Chan, D., Weng, Y. M., Hocking, A. M., Golub, S., McQuillan, D. J., and Bateman, J. F. (1996) *J.Biol.Chem.* **271,** 13566-13572

29. Sudo, H., Kodama, H., Amagai, Y., Yamamoto, S., and Kasai, S. (1983) *J.Cell Biol.* **96,** 191-198

30. Schnieke, A., Harbers, K., and Jaenisch, R. (1983) *Nature* **304,** 315-320

31. McMahon, D. K., Anderson, P. A., Nassar, R., Bunting, J. B., Saba, Z., Oakeley, A., and Malouf, N. N. (1994) *Am.J.Physiol.* **266,** 1795-1802

32. Chan, D., Taylor, T. K., and Cole, W. G. (1993) *J.Biol.Chem.* **268,** 15238-15245

33. Lefebvre, V., Garofalo, S., and de Crombrugghe, B. (1995) *J.Cell Biol.* **128,** 239-245

34. Chomczynski, P. and Sacchi, N. (1987) *Anal.Biochem.* **162,** 156-159

35. Kaelin, W. G. J., Krek, W., Sellers, W. R., DeCaprio, J. A., Ajchenbaum, F., Fuschs, C. S., Chittenden, T., Li, Y., Farnham, P. J., and Blanar, M. A. (1992) *Cell* **70,** 351-364

36. Paulsson, M. and Heinegard, D. (1982) *Biochem.J.* **207,** 207-213

37. Bateman, J. F., Mascara, T., Chan, D., and Cole, W. G. (1984) *Biochem.J.* **217,** 103-115

38. Nielson, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Engineering* **10,** 1-6

39. Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J. O., Hansen, J.-E. S., and Brunak, S. (1995) *Biochem.J.* **308,** 801-813

40. Leahy, D. J., Aukhil, I., and Erickson, H. P. (1996) *Cell* **84,** 155-164

41. Hauser, N., Paulsson, M., Heinegard, D., and Morgelin, M. (1996) *J.Biol.Chem.* **271,** 32247-32252

42. Murphy, J. M., Heinegard, D., McIntosh, A., Sterchi, D., and Barry, F. P. (1999) *Matrix Biol.* **18,** 487-497

43. Aszodi, A., Bateman, J. F., Hirsch, E., Baranyi, M., Hunziker, E. B., Hauser, N., Bosze, Z., and Fassler, R. (1999) *Mol.Cell Biol.* **19,** 7841-7845

44. Segat, D., Frie, C., Nitsche, D. P., Klatt, A. R., Piecha, D., Korpos, E., Deak, F., Wagener, R., Paulsson, M., and Smyth, N. (2000) *Matrix Biol.* **19,** 649-655

45. Ricard-Blum, S., Dublet, B., and van der Rest, M. (2000) *Unconventional collagens: types VI, VII, VIII, IX, X, XII, XIV, XVI and XIX*, 1st Ed., Oxford University Press, Oxford

46. Beck, K., Gambee, J. E., Bohan, C. A., and Bachinger, H. P. (1996) *J.Mol.Biol.* **256,** 909-923

47. Haudenschild, D. R., Tondravi, M. M., Hofer, U., Chen, Q., and Goetinck, P. F. (1995) *J.Biol.Chem.* **270,** 23150-23154

48. Schuppan, D., Cantaluppi, M. C., Becker, J., Veit, A., Bunte, T., Troyer, D., Schuppan, F., Schmid, M., Ackermann, R., and Hahn, E. G. (1990) *J.Biol.Chem.* **265,** 8823-8832

49. Young, B. B., Gordon, M. K., and Birk, D. (2000) *Dev.Dyn.* **217,** 439

50. Koch, M., Bohrmann, B., Matthison, M., Hagios, C., Trueb, B., and Chiquet, M. (1995) *J.Cell Biol.* **130,** 1005-1014

51. Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996) *Quart.Rev.Biophys.* **29,** 119-16

52. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucl.Acids Res.* **22,** 4673-468

53. Altschul *et al.*, (1997) *Nucl. Acids Res.* **25**:3389.

54. Ausubel *et al.*, "Current Protocols in Molecular Biology" John Wiley & Sons Inc, 1994-1998, Chapter 15

55. Bonner and Laskey (1974) *Eur. J. Biochem.* **46:**83Marmur and Doty (1962) *J. Mol. Biol.* **5:**10

56. Sambrook *et al.* (1990) Molecular Cloning – A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

SEQUENCE LISTING

<110> Murdoch Childrens Research Institute

<120> A molecular marker

<130> 2404275/EJH

<160> 19

<170> PatentIn version 3.0

<210> 1
<211> 537
<212> DNA
<213> human

<400> 1
ggggacctga tgttcctgct ggacagctca gccagcgtct ctcactacga gttctcccgg 60
gttcgggagt ttgtggggca gctggtggct ccactgcccc tgggcaccgg ggccctgcgt 120
gccagtctgg tgcacgtggg cagtcggcca tacaccgagt tccccttcgg ccagcacagc 180
tcgggtgagg ctgcccagga tgcggtgcgt gcttctgccc agcgcatggg tgacacccac 240
actggcctgg cgctggtcta tgccaaggaa cagctgtttg ctgaagcatc aggtgcccgg 300
ccaggggtgc ccaaagtgct ggtgtgggtg acagatggcg gctccagcga ccctgtgggc 360
ccccccatgc aggagctcaa ggacctgggc gtcaccgtgt tcattgtcag caccggccga 420
ggcaacttcc tggagctgtc agccgctgcc tcagcccctg ccgagaagca cctgcacttt 480
gtggacgtgg atgacctgca catcattgtc caagagctga ggggctccat tctcgcg 537

<210> 2
<211> 180
<212> PRT
<213> human

<400> 2

Arg Gly Asp Leu Met Phe Leu Leu Asp Ser Ser Ala Ser Val Ser His
1 5 10 15

Tyr Glu Phe Ser Arg Val Arg Glu Phe Val Gly Gln Leu Val Ala Pro
20 25 30

Leu Pro Leu Gly Thr Gly Ala Leu Arg Ala Ser Leu Val His Val Gly
35 40 45

Ser Arg Pro Tyr Thr Glu Phe Pro Phe Gly Gln His Ser Ser Gly Glu
50 55 60

Ala Ala Gln Asp Ala Val Arg Ala Ser Ala Gln Arg Met Gly Asp Thr
65 70 75 80

His Thr Gly Leu Ala Leu Val Tyr Ala Lys Glu Gln Leu Phe Ala Glu
85 90 95

Ala Ser Gly Ala Arg Pro Gly Val Pro Lys Val Leu Val Trp Val Thr
100 105 110

Asp Gly Gly Ser Ser Asp Pro Val Gly Pro Pro Met Gln Glu Leu Lys
115 120 125

Asp Leu Gly Val Thr Val Phe Ile Val Ser Thr Gly Arg Gly Asn Phe
130 135 140

Leu Glu Leu Ser Ala Ala Ala Ser Ala Pro Ala Glu Lys His Leu His
145 150 155 160

Phe Val Asp Val Asp Asp Leu His Ile Ile Val Gln Glu Leu Arg Gly
165 170 175

Ser Ile Leu Asp
180

<210> 3
<211> 1266
<212> DNA
<213> mouse

<400> 3
atgctgttct ggactgcgtt cagcatggct ttgagtctgc ggttggcatt ggcgcggagc 60
agcatagagc gcggttccac agcatcagac ccccagggg acctgttgtt cctgttggac 120
agctcagcca gcgtgtcaca ctatgagttc tcaagagttc gggaatttgt ggggcagctg 180
gtggctacga tgtctttcgg acccggggct ctgcgtgcta gtctggtgca cgtgggcagc 240
cagcctcaca cagagtttac ttttgaccag tacagttcag gccaggctat acgggatgcc 300
atccgtgttg caccccaacg tatgggtgat accaacacag gcctggcact ggcttatgcc 360
aaagaacaat tgtttgctga ggaagcaggt gcccggccag ggttcccaa ggtgctggtg 420
tgggtgacag atggtggctc cagcgacccc gtgggccccc ctatgcagga gctcaaggac 480
ctgggtgtca ccatcttcat tgtcagcact ggccgaggca acctgttgga gctgttggca 540

gctgcctcgg ctcctgccga gaagcaccta cactttgtgg atgtggatga tcttcctatc 600
attgcccggg agctgcgggg ctccataact gatgcgatgc agccacaaca gcttcatgcc 660
tcggaggttc tgtccagtgg cttccgcctg tcctggccgc ccctgctgac agcggactct 720
ggttactacg tgctggaatt ggtacctagc ggcaaactgg caaccacaag acgccaacag 780
ctgcccggga atgctaccag ctggacctgg acagatctcg acccggacac agactatgaa 840
gtatcactgc tgcctgagtc caacgtgcac ctcctgaggc cgcagcacgt gcgagtacgc 900
acactgcaag aggaggccgg gccagaacgc atcgtcatct cgcatgcgag gccgcgcagc 960
ctccgcgtaa gctgggcccc cgcgcttggc ccggactccg ctctcggcta ccatgtacag 1020
ctcggacctc tgcagggcgg gtccctagag cgcgtggagg tgccagcagg ccagaacagc 1080
actaccgtcc agggcctgac gccctgcacc acttacctgg tgactgtgac tgccgccttc 1140
cgctccggcc gccagagggc gctgtcggct aaggcctgta cggcctctgg cgcgcggacc 1200
cgtgctccgc agtccatgcg gccggaggct ggaccgcggg agccctgaac tgcctgcctg 1260
ctcgtc 1266

<210> 4
<211> 415
<212> PRT
<213> mouse

<400> 4

Met Leu Phe Trp Thr Ala Phe Ser Met Ala Leu Ser Leu Arg Leu Ala
1 5 10 15
Leu Ala Arg Ser Ser Ile Glu Arg Gly Ser Thr Ala Ser Asp Pro Gln
20 25 30
Gly Asp Leu Leu Phe Leu Leu Asp Ser Ser Ala Ser Val Ser His Tyr
35 40 45
Glu Phe Ser Arg Val Arg Glu Phe Val Gly Gln Leu Val Ala Thr Met
50 55 60
Ser Phe Gly Pro Gly Ala Leu Arg Ala Ser Leu Val His Val Gly Ser
65 70 75 80
Gln Pro His Thr Glu Phe Thr Phe Asp Gln Tyr Ser Ser Gly Gln Ala
85 90 95
Ile Arg Asp Ala Ile Arg Val Ala Pro Gln Arg Met Gly Asp Thr Asn
100 105 110
Thr Gly Leu Ala Leu Ala Tyr Ala Lys Glu Gln Leu Phe Ala Glu Glu
115 120 125
Ala Gly Ala Arg Pro Gly Val Pro Lys Val Leu Val Trp Val Thr Asp

130 135 140

Gly Gly Ser Ser Asp Pro Val Gly Pro Pro Met Gln Glu Leu Lys Asp

145 150 155 160

Leu Gly Val Thr Ile Phe Ile Val Ser Thr Gly Arg Gly Asn Leu Leu

165 170 175

Glu Leu Leu Ala Ala Ala Ser Ala Pro Ala Glu Lys His Leu His Phe

180 185 190

Val Asp Val Asp Asp Leu Pro Ile Ile Ala Arg Glu Leu Arg Gly Ser

195 200 205

Ile Thr Asp Ala Met Gln Pro Gln Gln Leu His Ala Ser Glu Val Leu

210 215 220

Ser Ser Gly Phe Arg Leu Ser Trp Pro Pro Leu Leu Thr Ala Asp Ser

225 230 235 240

Gly Tyr Tyr Val Leu Glu Leu Val Pro Ser Gly Lys Leu Ala Thr Thr

245 250 255

Arg Arg Gln Gln Leu Pro Gly Asn Ala Thr Ser Trp Thr Trp Thr Asp

260 265 270

Leu Asp Pro Asp Thr Asp Tyr Glu Val Ser Leu Leu Pro Glu Ser Asn

275 280 285

Val His Leu Leu Arg Pro Gln His Val Arg Val Arg Thr Leu Gln Glu

290 295 300

Glu Ala Gly Pro Glu Arg Ile Val Ile Ser His Ala Arg Pro Arg Ser

305 310 315 320

Leu Arg Val Ser Trp Ala Pro Ala Leu Gly Pro Asp Ser Ala Leu Gly

325 330 335

Tyr His Val Gln Leu Gly Pro Leu Gln Gly Gly Ser Leu Glu Arg Val

340 345 350

Glu Val Pro Ala Gly Gln Asn Ser Thr Thr Val Gln Gly Leu Thr Pro

355 360 365

Cys Thr Thr Tyr Leu Val Thr Val Thr Ala Ala Phe Arg Ser Gly Arg

370 375 380

Gln Arg Ala Leu Ser Ala Lys Ala Cys Thr Ala Ser Gly Ala Arg Thr

385 390 395 400

Arg Ala Pro Gln Ser Met Arg Pro Glu Ala Gly Pro Arg Glu Pro

405 410 415

<210> 5

<211> 1254

<212> DNA

<213> human

<400> 5

atgctcccct ggacggcgct cggcctggcc ctgagcttgc ggctggcgct ggcgcggagc 60
ggcgcggagc gcggtccacc agcatcagcc ccccgagggg acctgatgtt cctgctggac 120
agctcagcca gcgtctctca ctacgagttc tcccgggttc gggagtttgt ggggcagctg 180
gtggctccac tgcccctggg caccggggcc ctgcgtgcca gtctggtgca cgtgggcagt 240
cggccataca ccgagttccc cttcggccag cacagctcgg gtgaggctgc ccaggatgcg 300
gtgcgtgctt ctgcccagcg catgggtgac acccacactg gcctggcgct ggtctatgcc 360
aaggaacagc tgtttgctga agcatcaggt gcccggccag gggtgcccaa agtgctggtg 420
tgggtgacag atggcggctc cagcgaccct gtgggccccc ccatgcagga gctcaaggac 480
ctgggcgtca ccgtgttcat tgtcagcacc ggccgaggca acttcctgga gctgtcagcc 540
gctgcctcag cccctgccga gaagcacctg cactttgtgg acgtggatga cctgcacatc 600
attgtccaag agctgagggg ctccattctc gcgatgcggc cgcagcagct ccatgccacg 660
gagatcacgt ccagcggctt ccgcctggcc tggccacccc tgctgaccgc agactcgggc 720
tactatgtgc tggagctggt gcccagcgcc cagccggggg ctgcaagacg ccagcagctg 780
ccagggaacg ccacggactg gatctgggcc ggcctcgacc cggacacgga ctacgacgtg 840
gcgctagtgc ctgagtccaa cgtgcgcctc ctgaggcccc agatcctgcg ggtgcgcacg 900
cggccagagg aggccgggcc agagcgcatc gtcatctccc acgcccggcc gcgcagcctc 960
cgcgtgagtt gggccccagc gctgggctca gccgcggcgc tcggctacca cgtgcagttc 1020
gggccgctgc ggggcgggga ggcgcagcgg gtggaggtgc ccgcgggccg caactgcacc 1080
acgctgcagg gcctggcgcc gggcaccgcc tacctggtga ccgtgaccgc cgccttccgc 1140
tcgggccgcg agagcgcgct gtccgccaag gcctgcacgc ccgacggccc gcgcccgcgc 1200
ccacgccccg tgccccgcgc cccgaccccg gggaccgcca gccgtgagcc gtaa 1254

<210> 6

<211> 418

<212> PRT

<213> human

<400> 6

Met Leu Pro Trp Thr Ala Leu Gly Leu Ala Leu Ser Leu Arg Leu Ala
1   5   10   15

Leu Ala Arg Ser Gly Ala Glu Arg Gly Pro Pro Ala Ser Ala Pro Arg
            20   25   30

Gly Asp Leu Met Phe Leu Leu Asp Ser Ser Ala Ser Val Ser His Tyr

```
        35                  40                  45
Glu Phe Ser Arg Val Arg Glu Phe Val Gly Gln Leu Val Ala Pro Leu
    50                  55                  60
Pro Leu Gly Thr Gly Ala Leu Arg Ala Ser Leu Val His Val Gly Ser
65                  70                  75                  80
Arg Pro Tyr Thr Glu Phe Pro Phe Gly Gln His Ser Ser Gly Glu Ala
                85                  90                  95
Ala Gln Asp Ala Val Arg Ala Ser Ala Gln Arg Met Gly Asp Thr His
            100                 105                 110
Thr Gly Leu Ala Leu Val Tyr Ala Lys Glu Gln Leu Phe Ala Glu Ala
        115                 120                 125
Ser Gly Ala Arg Pro Gly Val Pro Lys Val Leu Val Trp Val Thr Asp
    130                 135                 140
Gly Gly Ser Ser Asp Pro Val Gly Pro Pro Met Gln Glu Leu Lys Asp
145                 150                 155                 160
Leu Gly Val Thr Val Phe Ile Val Ser Thr Gly Arg Gly Asn Phe Leu
                165                 170                 175
Glu Leu Ser Ala Ala Ala Ser Ala Pro Ala Glu Lys His Leu His Phe
            180                 185                 190
Val Asp Val Asp Asp Leu His Ile Ile Val Gln Glu Leu Arg Gly Ser
        195                 200                 205
Ile Leu Asp Ala Met Arg Pro Gln Gln Leu His Ala Thr Glu Ile Thr
    210                 215                 220
Ser Ser Gly Phe Arg Leu Ala Trp Pro Pro Leu Leu Thr Ala Asp Ser
225                 230                 235                 240
Gly Tyr Tyr Val Leu Glu Leu Val Pro Ser Ala Gln Pro Gly Ala Ala
                245                 250                 255
Arg Arg Gln Gln Leu Pro Gly Asn Ala Thr Asp Trp Ile Trp Ala Gly
            260                 265                 270
Leu Asp Pro Asp Thr Asp Tyr Asp Val Ala Leu Val Pro Glu Ser Asn
        275                 280                 285
Val Arg Leu Leu Arg Pro Gln Ile Leu Arg Val Arg Thr Arg Pro Glu
    290                 295                 300
Glu Ala Gly Pro Glu Arg Ile Val Ile Ser His Ala Arg Pro Arg Ser
305                 310                 315                 320
Leu Arg Val Ser Trp Ala Pro Ala Leu Gly Ser Ala Ala Ala Leu Gly
                325                 330                 335
Tyr His Val Gln Phe Gly Pro Leu Arg Gly Gly Glu Ala Gln Arg Val
```

```
                340                 345                 350
Glu Val Pro Ala Gly Arg Asn Cys Thr Thr Leu Gln Gly Leu Ala Pro
            355                 360                 365
Gly Thr Ala Tyr Leu Val Thr Val Thr Ala Ala Phe Arg Ser Gly Arg
        370                 375                 380
Glu Ser Ala Leu Ser Ala Lys Ala Cys Thr Pro Asp Gly Pro Arg Pro
385                 390                 395                 400
Arg Pro Arg Pro Val Pro Arg Ala Pro Thr Pro Gly Thr Ala Ser Arg
                    405                 410                 415
Glu Pro


<210>  7
<211>  539
<212>  DNA
<213>  mouse

<400>  7
aggggacct gttgttcctg ttggacagct cagccagcgt gtcacactat gagttctcaa      60
gagttcggga atttgtgggg cagctggtgg ctacgatgtc tttcggaccc ggggctctgc     120
gtgctagtct ggtgcacgtg ggcagccagc ctcacacaga gtttactttt gaccagtaca     180
gttcaggcca ggctatacgg gatgccatcc gtgttgcacc ccaacgtatg ggtgatacca     240
acacaggcct ggcactggct tatgccaaag aacaattgtt tgctgaggaa gcaggtgccc     300
ggccaggggt tcccaaggtg ctggtgtggg tgacagatgg tggctccagc gaccccgtgg     360
gcccccctat gcaggagctc aaggacctgg gtgtcaccat cttcattgtc agcactggcc     420
gaggcaacct gttggagctg ttggcagctg cctcggctcc tgccgagaag cacctacact     480
ttgtggatgt ggatgatctt cctatcattg cccgggagct gcggggctcc ataactgat      539


<210>  8
<211>  180
<212>  PRT
<213>  mouse

<400>  8
Gln Gly Asp Leu Leu Phe Leu Leu Asp Ser Ser Ala Ser Val Ser His
1               5                   10                  15
Tyr Glu Phe Ser Arg Val Arg Glu Phe Val Gly Gln Leu Val Ala Thr
            20                  25                  30
Met Ser Phe Gly Pro Gly Ala Leu Arg Ala Ser Leu Val His Val Gly
```

35 40 45

Ser Gln Pro His Thr Glu Phe Thr Phe Asp Gln Tyr Ser Ser Gly Gln

50 55 60

Ala Ile Arg Asp Ala Ile Arg Val Ala Pro Gln Arg Met Gly Asp Thr

65 70 75 80

Asn Thr Gly Leu Ala Leu Ala Tyr Ala Lys Glu Gln Leu Phe Ala Glu

85 90 95

Glu Ala Gly Ala Arg Pro Gly Val Pro Lys Val Leu Val Trp Val Thr

100 105 110

Asp Gly Gly Ser Ser Asp Pro Val Gly Pro Pro Met Gln Glu Leu Lys

115 120 125

Asp Leu Gly Val Thr Ile Phe Ile Val Ser Thr Gly Arg Gly Asn Leu

130 135 140

Leu Glu Leu Leu Ala Ala Ala Ser Ala Pro Ala Glu Lys His Leu His

145 150 155 160

Phe Val Asp Val Asp Asp Leu Pro Ile Ile Ala Arg Glu Leu Arg Gly

165 170 175

Ser Ile Thr Asp

180

<210> 9
<211> 20
<212> DNA
<213> primer

<400> 9
ctcaaagcca tgcgtagtcc 20

<210> 10
<211> 20
<212> DNA
<213> primer

<400> 10
agaacgcatc gtcatctcgc 20

<210> 11
<211> 20

<212> DNA
<213> primer

<400> 11
agaacgcatc gtcatctcgc 20

<210> 12
<211> 20
<212> DNA
<213> primer

<400> 12
tcaagggcat atccaacaac 20

<210> 13
<211> 20
<212> DNA
<213> primer

<400> 13
ctggtcatcg ccgcccttgc 20

<210> 14
<211> 22
<212> DNA
<213> primer

<400> 14
gaccagcgtt aattcctttc gt 22

<210> 15
<211> 17
<212> DNA
<213> primer

<400> 15
ccgggtttcc cggaagt 17

<210> 16
<211> 32
<212> DNA
<213> primer

<400> 16
ttactggcaa catcaacagg actcctcgta tt 32

<210> 17
<211> 24
<212> DNA
<213> primer

<400> 17
ccacaggact agaacacctg ctaa 24

<210> 18
<211> 22
<212> DNA
<213> primer

<400> 18
cctaagatga gcgcaagttg aa 22

<210> 19
<211> 9060
<212> DNA
<213> human

<400> 19
cctctgcatt ccagccacct gccctgggcc cagctccaaa ggaagggggc ccaagctctc 60
tgaataaaag gtgcacatga ggaccaagga ggcctgacac tgggaggga cagctccacc 120
tcctctcccc ggacacccca aaaggcggag acgttcacaa gctgtcctgt cggcggctgc 180
tgtttgtgga ggagtaaagc atcctagcga gactgcaggc tcggtgtaca tctgatttac 240
tgaattttaa agtctgggat gttagtgggg aagaggcgag gtgagcattg cgtgacgccg 300
aggactaggc ggggcgggga ctgcacctgg ctaggcaccc ccaccctggg caacttgccc 360
acggacccca gggcagtgag tagtgacagg aggtagcccg gggtgagacc tctcacagca 420
agaagatggt gtggttgctg gggcctccct ggagagtgtc gtccctgcgg cccctgggaa 480

gtgctccctc acgacggaag gtttcctgtc agtgcggtcc cggggcctga tagtggcggt 540
gggcgggtgg ggtcacgtgt cctcaaggtc ctgaatgccc agctctgccc cattcctctg 600
attcccagtg gctgctagct ggacccagct ggtgtcctgg gcatgaaggc agggccaccg 660
tccccagcag gtgctgccct cctggccagc tgagcatcct ggccaccatc agcgtccagg 720
tgcccctact cgcccttcct cttcttcaga agcctttgcg gacctgacct gggccagctt 780
cccgcgattc cccttccgct tcctatcaac gtccaggacc caagctgccc gccccaggcc 840
agcccttgcc acttggggcc cggtcttcac acgtgggagt ctgaccgggg ctcctccctg 900
aacagtcctg ggtctgacgc tctcaattat cacccacgga cccacacgac gcccggctct 960
gggcggggat ggggccgggg ctgctgcggg gtcccgccag gcgaggcccc agccctggag 1020
ggcaggcgcc aggcggggaa gccctgcggc cgcagggaga gggccggggt cgcgcggagt 1080
ccgcgtgggg aaaggccggg cctgcacccg tctgccgggt tgggcgcctc cgctccgggt 1140
tcgggacaca ggggccctca ggtaggcgcc ggccctctcg gctgggcggg gacgccggct 1200
tacggctcac ggctggcggt ccccggggtc ggggcgcggg gcacggggcg tgggcgcggg 1260
cgcgggccgt cgggcgtgca ggccttggcg gacagcgcgc tctcgcggcc cgagcggaag 1320
gcggcggtca cggtcaccag gtaggcggtg cccggcgcca ggccctgcag cgtggtgcag 1380
ttgcggcccg cgggcacctc cacccgctgc gcctccccgc cccgcagcgg cccgaactgc 1440
acgtggtagc cgagcgccgc ggctgagccc agcgctgggg cccaactcac gcggaggctg 1500
cgcggccggg cgtgggagat gacgatgcgc tctggcccgg cctcctctgg ggcggggagg 1560
gcggcgagct gcgtgggggc cggcccagcc cccgactccg ggcccgaagc ccccggccct 1620
gcctcaccgg gccgcgtgcg cacccgcagg atctggggcc tcaggaggcg cacgttggac 1680
tcaggcacta gcgccacgtc gtagtccgtg tccgggtcga ggccggccca gatccagtcc 1740
gtggcgttcc ctggcagctg ctggcgtctt gcagcccccg gctgggcgct gggcaccagc 1800
tccagcacat agtagcccga gtctgcggtc agcaggggtg gccaggccag gcggaagccg 1860
ctggacgtga tctccgtggc atggagctgc tgcggccgca tcgcgtctgt gggtggtgca 1920
gggggtcagg gaacagcggt cagttcctcc tccgctgctg gagggcggcc ctggctgatg 1980
gggaagatct ggagattgga ggccccacta ggaaagacgg ggccccgcgg ccaaggagct 2040
gctggagcca tgccccgcag atgctgggga ttctcagaac gtgccttggc tgggggagga 2100
cggaggaaag ggtgcagccc cctcaggccc tgtcagaagc gcccctgcct cccttagccc 2160
caaacccagt cctttgtgga gaggtgcagt ggccagatca gtgaccagga caaaggtcct 2220
caaagacggc agagtccacg gtggtgcctg agagcagagg accagcccca gcctgagtgg 2280
ccaggccggg gtctgaggtc agcccggctc tctgagctgc agctaggaga tgggagacca 2340
caggggcagg ccctggggtt ctggaggcgc tgcctgccct gggtccccag gagagtgtgg 2400
ggtggggttc tccagagggg gactcctgga cctgtgacac caagccccac atagccctct 2460
gagtgaccct gctgtggcga ggctcataaa tgtctgcgct gggttaaagc tatcaggatc 2520
ttcctcctgc agtgctgggt gcctgggcca ctttcttccc atcccccacc ctcagacccg 2580
gcctctttcc caggagcccc caccctgctg cctggcccct cggcactgca gcctcaggct 2640
tttcctttgg ctgcttaagg cagcctttcc tcctggtccc ctccaggcgc agctgcactg 2700
ggtgacctgg ggccactagg ggccagacgt ccctggggaa accttgggga gggccgtcca 2760

```
ccccctctcca acccacagtc caacccccttc cggctctggg tggatgatta acccacagac    2820
ggagacttgg tgagatcccc agggttggca tttttcagtg gctgcagcag gctgagccag    2880
tggccggttc ctcatctcca gccccagctc cttcagggct tggctggggc agggaggtcc    2940
agaaaaaaag ccaatgggag ctgctcagct cctgcctcag gccttccctg gtccggcctc    3000
tcaggaaacc ctcacagtgg gcctgcagtc cgaactagtt caaagccctc ggcggctgtc    3060
cccacccagg agaggtgccc tgtgctctct gggggggcag tccctgacct ttctggctca    3120
cccctctcca ggtatggtgg gcatgctcag gagcacatgc tgcccatctg cagagtcccc    3180
agacttggaa gcttcttcct gggcctacac ccgggctctg cactccctgg ggcctcgagg    3240
tctgggctgg acacatcagc agggagctac acctggaggt ggctactcaa gcctgccccc    3300
gtctcagcag ggtacacggg tcgcccagtg aagagtgtgc atagacaagc tgcatcactc    3360
agccctgcac cctaggggta ccacagcccc ggaggccctg gccgctgctc tggggacatg    3420
agatcttccc aaagtctcaa cccagcctct ccttctgcgg ctcccagcta gggctccctg    3480
ggccctgcct cctcccgcat accgagaatg gagcccctca gctcttggac aatgatgtgc    3540
aggtcatcca cgtccacaaa gtgcaggtgc ttctcggcag ggctgaggc agcggctgac    3600
agctccagga agttgcctcg gccggtgctg acaatgaaca cggtgacgcc caggtccttg    3660
agctcctgca tgggggggcc cacagggtcg ctggagccgc catctgtcac ccacaccagc    3720
actttgggca cccctggccg ggcacctgat gcttcagcaa acagctgttc cttggcatag    3780
accagcgcca ggccagtgtg ggtgtcaccc atgcgctggg cagaagcacg caccgcatcc    3840
tgggcagcct cacccgagct gtgctggccg aaggggaact cggtgtatgg ccgactgccc    3900
acgtgcacca gactggcacg cagggccccg gtgcccaggg gcagtggagc caccagctgc    3960
cccacaaact cccgaacccg ggagaactcg tagtgagaga cgctggctga gctgtccagc    4020
aggaacatca ggtcccctcg gggggctgat gctggtggac ctgggggaaa ggaggaatgc    4080
tcagcctcag gtgtgggccc cccagacagc cccacagcaa ggcagggtcc cccagggccc    4140
cagctttcct taagtggatg cttgccttct cccaaaggtc ctaggttggg ggaaagagga    4200
actctaagca agaggcctgt acttttgggg gtttcactgc acactggcca tgggatctag    4260
ggctctctct gggcttgtgt tatcccatct gtgagagggc gactctccgc tccaagcccc    4320
cacaccttcc cattcctcac agaccctgca agcaggtgga gccaagagtc ctggcctagg    4380
cccccaggac aggcctgagc cgtggggctg ttccctccag gcatggcttt cagaggagca    4440
gcctgaggct ggagttcagc cacgcagctc agcctgcagg tgaggcaccc tgggcatgca    4500
cacagcagca ggggaaggtg tcggaggcac agcaatgacc acgccggatg gcctggctgg    4560
agcccagacc ccgcttacta gatggtggcc cctcccctgg cctccatcct ccagcccacc    4620
tggactcaca caacaagata taaccccccag cagcctgaaa gccggaacag cccctcgcag    4680
gcttcccccct tcctccgggc acctccgggg tggaggctga tgccccctac accgcccctc    4740
cccaccaagc cagggcacca gcgtgcctca attctagtcc cggccttgcg gttttcccca    4800
gtgcggtggg gcgactccaa cttccctacc atccctccac taagggccct cgcaagggta    4860
gggaaactga ggcaggggtg ccccccttgac agacatctcc ctcttcctgt ccaggcccgc    4920
gatcccgcag agatgcgggc cgggacggcc cctatgcccc ggcgctcacg gacggtgtcg    4980
cctggagcac ctgggccgcc agcctcaggt gagcaggacg ctccgcccgc gcccccgccc    5040
```

ggctcccgca gcctcccagc ccgcccgccc gtccggagca ggggacagcg acggccttgc 5100
gcgggcagcg gcgcagagcg gtcaccagaa gccccagccc cggcccggcc gcccgccgca 5160
ctcaccgcgc tccgcgccgc tccgcgccag cgccagccgc aagctcaggg ccaggccgag 5220
cgccgtccag gggagcatcg cgcgcgaggg acggggcgcg ctcggcaact cgctcgctcg 5280
ctcgctcgct cggggctgca gggcgcgtca ccgcgcggac caggccggcc ccgcccccgg 5340
gaggcccctc cccgagcggc cacacccacg ccgaggccac gcccacgccc tccggcgcga 5400
gcggagggcc acgcgcacag accccggaga ggcgcgcacg agcggacccc gacacgcagg 5460
gacacgcagc accagccgag atacgaccga ggcacgcacg cgcaggcacg cacacacaca 5520
cactccagtc tccctctccc ggccgaggct gtgcggccca cgctctccac ccctctccga 5580
cccccagccg cgggagccga gcagggaggt accaggctag gccctcccca tgcccaccac 5640
tgccgtgact ctgggtgctg gggtcccagc agccaggccc aagagaaccc caggggctgg 5700
cggtggcacc aaaaaaacac gtccagaccg tggtttcgcc ttggcctccg cgctggaggc 5760
ggataggtgt ctggagtaac aggacatgta tcccagggac tgaccagcag ggatgggaag 5820
gaccatgggg tggaacttac aaggacacag tggcttgaaa ggggacagaa gacaggaatt 5880
cgagagagac tcgaagcacc cacgccacct gggcttcttg gaggaagagg catgggagtg 5940
ggagatggtt ggttgaggcc ctgtccagtg ggaccacact gggcctgtta cccatatacc 6000
ctacccagtg aggggcccag actccaggac ccaggacaca cccccagcag gactggaggg 6060
tcccactggt gagacaggag ctcttgagtc ttggggtctt ggtgaggccc agacgagagg 6120
tggctggttg caggggggcgt cctgagggac agtggctccc agggcagatt tcccctgctt 6180
gggtggggct gggccagcag tgtcccctgg acaggagaac cctaccccgg ccctccctcg 6240
gagtagccat ggccctcttc cagggcctcc tcagctcaga gctgggaggt gggggacgtg 6300
ggggggtgtc tgccaggatg tctcctcctt ccccaccctc tcctggagga tgcgccgcgg 6360
gagaacggat ggggctccac aggcttcctt cctccctttc aggcaggtga gacaccgcgg 6420
ggccgtgcgg acggccagca ctcgactttg cctaaaaaag gaagcagcag gctgaggctg 6480
aggagctggc ggcaggaaca agggagagct gtgtccccgc cggcgccccc cacccccccct 6540
gccggggatc ttggcagtgg aggtgctggc tgcgctccac agacctcaga cctcggctgg 6600
gaccagaaat gcctggtgct tccgcctggg cccggtgggg ggactttggg tccccagagt 6660
gcaagctgta ccacttcgag ggcctcgcc aggcccccca gcccccagta cacaggggct 6720
gccgtggaga tgacgctgaa ggccgcagcc gctggaggac ctggggtctg accggaagct 6780
ggctgcagac cctgcggagg cacgtccagg tagtcaggca gggagctggg ccgagggtcc 6840
cccaccctgg ggaggctcac agccagtggc ccgcttgtcc cccaccctcg cccagcaggc 6900
gggccacagt cacacctcag ccagccttgc agggctgacg gggaagtttc cctcacttct 6960
ggaaaaagtg agcgggtctt cttggctgtg actcaggccc tcaaggaagc ggccgccctc 7020
ctcccttcag ctcgccatca gcgggagaag gcacaggagg cctggcctcc acccagcctg 7080
ggccgagctc agccacctgc cttgctcccg gctctgcctg gagtccctcc agctaggaga 7140
ccctccccat cagctctccc cgtgcccctc agtcttcagg actcattctt gtgtcctgcc 7200
ctccccccgc tgtctccacc ccggaggagg gacgtggaca gagggtccca gagagcatgg 7260
ggtcagccag aggtgcagtg tcagggcccg ggccggactt gaggcagaca ccggaggaag 7320

cacaaatata acagccggaa ccctccactc tccagggaga agggcccggg gtaagaggca 7380
gaggcaagga cgggtcaggc cagatcacag tgggtgctgg ccccgagccc tctgcctcct 7440
gcaggcacag cccctgtctg atcctggtgg cctggggccc catggggtgg ggagcagcct 7500
ggtttggctg cggccacccc gcccccacgg tctgggcctg ggctgtggga gtccctgtgc 7560
ctcacttccc ggagccagcc tgccctgccg gtctgtctgc aggcaggtgg agagagttcc 7620
aggaagctgg ggaggctgct gtcacccggg caccgcccct gcccccaccc gcctttggga 7680
atgctccctc ctccgcacaa tccaggcttc tgcagaagat gaagggcctt ttgtccccag 7740
ctggctgtgg tcatgtttga ccctgggtaa aagggcaact cctgaggcct ctgaccccac 7800
ccctgacccg agctgagggc aggacgccca ggcccgcacc cggcgccttt tgttgctgtt 7860
ttcacgtatc tcacaaacgt actcaagcac acacaggagc agatggacgg ggcggtgagg 7920
ggcagcagtg gtgaggggca gcggcggtga ggggcagcgg cggtgagggg cagcggtgcg 7980
ggcctgaggc actgctctgg ggtgtgcctg agcccacccc acaacagtaa gtggggcaga 8040
gcaggggtca ccaagagagc agggcccacg cagctcctag actcaacctg ctcactgggg 8100
tcaaggacag gtcttggggg cctcgggggt cacttttcac ttcccaggag cccaggcctg 8160
cccctctggc cccagagctg acccccctca gtcccccgtg ccagcagcag ctggggtggc 8220
gggtagacac ctggcgggta gcagcctggg taggggtggg agctgcacca tctgcgtctg 8280
tccatccatc cctcgtctgt gtgctgggca cagccgcgcc ccagcctcag tgctggggac 8340
acacaggcgc cgggccagca ctgccaggct aggagggtgg gcggtgaaca gctaggaaag 8400
atacggtcta cttgttttcc ctgtgagaac aggggtcac tggggactcg cacgcaaggg 8460
gtacccgagg aagagccttc caggcagaga gaaggaaccg cgagtgctga gagcagggtg 8520
gggtgggcag gaggggcctg cgccaggact gcaggggcag agcaggctgg gggccttcgg 8580
gaggggtggc cgggtggagg gtgttgccgg cctcgacagg ggcaggaggt tcgtcacagc 8640
gaggacagag cccggcccgg tgggagccgg agagcagcag gcctgaatga cccagggttt 8700
cctaatagca gggccccttc cttgtgtggg tcccctcact ttgcctctct gctgggacat 8760
ccttccctga aagggagagg aggaccacat gctgcccctt ccccagacac agtccagaca 8820
ggcccaggcc acagccctgg gcagacgcaa aactcccagg ggcctggact gggataggga 8880
ggaggcagca gggagggact gacctatgtc cacacaccac aagggactcc cagaggcggg 8940
tggggcggag ctgggagcag gggccttagc cctcagacca gcccactcac cctggggagt 9000
tcctgcccca cagcctgccc agcttacagg cctgggggca ggggcaggcc agcacaggcc 9060

DATED this second day of May 2001.

Murdoch Childrens Research Institute
by DAVIES COLLISION CAVE
Patent Attorneys for the Applicant

```
tcgatcaagagcccgccactccaggcgcgatgctgttctggactgcgttcagcatggctttgagtctgcgg   71
                               M  L  F  W  T  A  F  S  M  A  L  S  L  R    14
ttggcattggcgcggagcagcatagagcgcggttccacagcatcagacccccaggggacctgttgttcctg  143
 L  A  L  AΔR  S  S  I  E  R  G  S  T  A  S  D  P  Q  G  D  L  L  F  L   38
ttggacagctcagccagcgtgtcacactatgagttctcaagagttcgggaatttgtggggcagctggtggct  215
 L  D  S  S  A  S  V  S  H  Y  E  F  S  R  V  R  E  F  V  G  Q  L  V  A   62
acgatgtctttcggacccggggctctgcgtgctagtctggtgcacgtgggcagccagcctcacacagagttt  287
 T  M  S  F  G  P  G  A  L  R  A  S  L  V  H  V  G  S  Q  P  H  T  E  F   86
acttttgaccagtacagttcaggccaggctatacgggatgccatccgtgttgcaccccaacgtatgggtgat  359
 T  F  D  Q  Y  S  S  G  Q  A  I  R  D  A  I  R  V  A  P  Q  R  M  G  D  110
accaacacaggcctggcactggcttatgccaaagaacaattgtttgctgaggaagcaggtgcccggccaggg  431
 T  N  T  G  L  A  L  A  Y  A  K  E  Q  L  F  A  E  E  A  G  A  R  P  G  134
gttcccaaggtgctggtgtgggtgacagatggtggctccagcgaccccgtgggcccccctatgcaggagctc  503
 V  P  K  V  L  V  W  V  T  D  G  G  S  S  D  P  V  G  P  P  M  Q  E  L  158
aaggacctgggtgtcaccatcttcattgtcagcactggccgaggcaacctgttggagctgttggcagctgcc  575
 K  D  L  G  V  T  I  F  I  V  S  T  G  R  G  N  L  L  E  L  L  A  A  A  182
tcggctcctgccgagaagcacctacactttgtggatgtggatgatcttcctatcattgcccgggagctgcgg  647
 S  A  P  A  E  K  H  L  H  F  V  D  V  D  D  L  P  I  I  A  R  E  L  R  206
ggctccataactgatgcgatgcagccacaacagcttcatgcctcggaggttctgtccagtggcttccgcctg  719
 G  S  I  T  D  A  M  Q  P  Q  Q  L  H  A  S  E  V  L  S  S  G  F  R  L  230
tcctggccgcccctgctgacagcggactctggttactacgtgctggaattggtacctagcggcaaactggca  791
 S  W  P  P  L  L  T  A  D  S  G  Y  Y  V  L  E  L  V  P  S  G  K  L  A  254
accacaagacgccaacagctgcccgggaatgctaccagctggacctggacagatctcgacccggacacagac  863
 T  T  R  R  Q  Q  L  P  G  N  A  T  S  W  T  W  T  D  L  D  P  D  T  D  278
tatgaagtatcactgctgcctgagtccaacgtgcacctcctgaggccgcagcacgtgcgagtacgcacactg  935
 Y  E  V  S  L  L  P  E  S  N  V  H  L  L  R  P  Q  H  V  R  V  R  T  L  302
caagaggaggccgggccagaacgcatcgtcatctcgcatgcgaggccgcgcagcctccgcgtaagctgggcc 1007
 Q  E  E  A  G  P  E  R  I  V  I  S  H  A  R  P  R  S  L  R  V  S  W  A  326
cccgcgcttggcccggactccgctctcggctaccatgtacagctcggacctctgcagggcgggtccctagag 1079
 P  A  L  G  P  D  S  A  L  G  Y  H  V  Q  L  G  P  L  Q  G  G  S  L  E  350
cgcgtggaggtgccagcaggccagaacagcactaccgtccagggcctgacgccctgcaccacttacctggtg 1151
 R  V  E  V  P  A  G  Q  N  S  T  T  V  Q  G  L  T  P  C  T  T  Y  L  V  374
actgtgactgccgccttccgctccggccgccagagggcgctgtcggctaaggcctgtacggcctctggcgcg 1223
 T  V  T  A  A  F  R  S  G  R  Q  R  A  L  S  A  K  A  C  T  A  S  G  A  398
cggacccgtgctccgcagtccatgcggccggaggctggaccgcgggagccctgaactgcctgcctgctcgtc 1295
 R  T  R  A  P  Q  S  M  R  P  E  A  G  P  R  E  P  *                    415
cacccgggggccctcttccctagcccggagagagagacactgctgctcgtgggtttcttgtggatggagtc 1367
gggtggggagatggatgccggtcctgcctttgaccagcgttaattcctttcgtcgtttccccactggtcat 1439
cgccgcccttgcctgacttccgggaaacccgggtagcctcacgcgcaatggcggtcctctccggttgccagt 1511
ggagttgagcacacggtggtccttgggcaactcttggcgaggggatggacagtgtctgaggtcaggttgagg 1583
acataagacccaggaaccgccttcaggagaggaggccacagagtttccaacctgtgccaaaggctgggccct 1655
ctggtggcagggactacgcatggctttgaggaggcgttcaggaccatccaggtcctgcctgggcctagaaag 1727
tgggtaggagaaagggaagagagactagtgtagacaggattcccgaaaacttcctcaaggaaaggaaagata 1799
gggaggtatgctgggaggctgatgatgtggcattggttttcatcaagatgtcctgccagcctagaggccggg 1871
atctgtcagggtcactgactctgccttcctgcccaggacctgcactgggccctcgatcagtgccaaggatgc 1943
agtcttttcacaggaatgggacgagaccttggcatttagggcctcagggataggagagccgcactatgacag 2015
attctaagggagcctcctgctttagtgtagggagcaaggtgtcatgcaggtgggctacctcctgccatcacc 2087
attaccctggggcatctgacagatacctaagggtggtcaggaacaggtttcctctcaagtccctatgtaggc 2159
ctctcctctcctctcagaatcatttgccttatcccaagcttactccatctcttccccactaatgacccggac 2231
tctaacaacaatacagtcagacagacataaactgtgcctgcagtctcattaaaatgctgtatttttcgtcaa 2303
aaaaaaaa                                                                 2308
```

Figure 1A

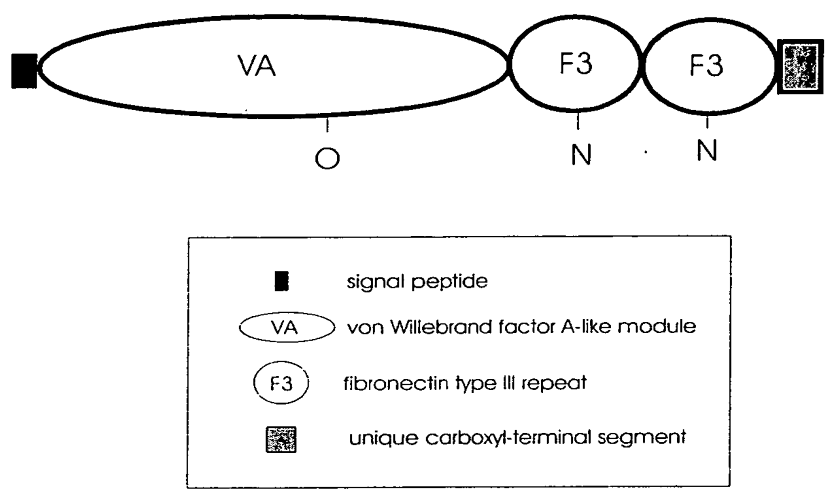

VA
F3
F3
O
N
N
signal peptide
VA
von Willebrand factor A-like module
F3
fibronectin type III repeat
unique carboxyl-terminal segment

Figure 1B

```
                         * * *
mat2 VA-1(m)       RADLVFLLDSSRSVNTYDYAKVKEFILDILQFL--DIGPDVTRVGLLQYGSTVKNEFSLK
mat4 VA-1(m)       PLDLVFMIDSSRSVRPFEFETMRQFLVGLLRSL--DVGLNATRVGVIQYSSQVQSVFPLG
mat3 VA-1(m)       PLDLVFLLDSSRSVRPLEFTKVKTFVSRIIDTL--DIGATDTRVAVVNYASTVKIEFQLN
mat1 VA-1(m)       PTDLVFVVDSSRSVRPVEFEKVKVFLSQVIESL--DVGPNATRVGLVNYASTVKPEFPLR
collXIV VA-1(ch)   IADLVILVDGSWSIGRFNFRLVRLFLENLVSAF--NVGSEKTRVGLAQYSGDPRIEWHLN
collXII VA-4(h)    KADLVFLTDASWSIGDDNFNKVVKFIFNTVGGFD-EISPAGIQVSFVQYSDEVKSEFKLN
collVII VA(h)      AADLVFLLDGSSSIGRSNFREVRSFLEGLVLPFSGAASAQGVRFATVQYSDDPRTEFGLD
collVI VA-1(m)     AADLVFLVDSSWSAGKDRFLLVQEFLSDVVESL--AVGDNDFHFALVRLNGNPHTEFLLN
WARP VA(m)         QGDLLFLLDSSASVSHYEFSRVREFVGQLVATM--SFGPGALRASLVHVGSQPHTEFTFD
cochlin VA-1(ch)   KADIAFLIDGSYNIGQRRFNLQKNFVGKVAVML--GIGTEGPHVGVVQASEHPKIEFYLK
VLA VA(h)          QLDIVIVLDGSNSIYP--WDSVTAFLNDLLKRM--DIGPKQTQVGIVQYGENVTHEFNLN
vWF VA-A1(h)       LLDLVFLLDGSSRLSEAEFEVLKAFVVDMMERL--RISQKWVRVAVVEYHDGSHAYIGLK

                                                *                               *
mat2 VA-1(m)       TFKRKSEVERAVKRMRHLST-GTMTGLAIQYALNIAFSEAEGARPLRENVPRIIMIVTDG
mat4 VA-1(m)       AFSRREDMERAIRAVVPLAQ-GTMTGLAIQYAMNVAFSEAEGARPSEERVPRVLVIVTDG
mat3 VA-1(m)       TYSDKQALKQAVARITPLST-GTMSGLAIQTAMEEAFTVEAGARGPMSNIPKVAIIVTDG
mat1 VA-1(m)       AHGSKASLLQAVRRIQPLST-GTMTGLALQFAITKALSDAEGGRARSPDISKVVIVVTDG
collXIV VA-1(ch)   AYGTKDAVLDAVRNLPYKGG-NTLTGLALTYILENSFKPEAGAR---PGVSKIGILITDG
collXII VA-4(h)    TYNDKALALGALQNIRYRGG-NTRTGKALTFIKEKVLTWESGMR---KNVPKVLVVVTDG
collVII VA(h)      ALGSGGDVIRAIRELSYKGG-NTRTGAAILHVADHVFLP-QLAR---PGVPKVCILITDG
collVI VA-1(m)     TYHSKQEVLSHIANMSYIGG-SNQTGKGLEYVIHSHLTEASGSR-AADGVPQVIVVLTDG
WARP VA(m)         QYSSGQAIRDAIRVAPQRMG-DTNTGLALAYAKEQLFAEEAGAR---PGVPKVLVWVTDG
cochlin VA-1(ch)   NFTAAKEVIFAIKELGFRGG-NSNTGKALKHAAQKFFSMENGAR---KGIPKIIVVFLDG
VLA VA(h)          KYSSTEEVLVAAKKIVQRGGRQTMTALGTDTARKEAFTEARGAR---RGVKKVMVIVTDG
vWF VA-A1(h)       DRKRPSELRRIASQVKYAGSQVASTSEVLKYTLFQIFSK--IDR---PEASRIALLLMAS

mat2 VA-1(m)       RPQD-----SVAEVAAKARNTGILIFAIGVGQVD-----------LNTLKAIGSE-PHKDH
mat4 VA-1(m)       RPQD-----RVAEVAAQARARGIEIYAVGVQRAD-----------VGSLRTMASP-PLDQH
mat3 VA-1(m)       RPQD-----QVNEVAARARASGIELYAVGVDRAD-----------MESLKMMASK-PLEEH
mat1 VA-1(m)       RPQD-----SVRDVSERARASGIELFAIGLGRVD-----------KATLRQIASE-PQDEH
collXIV VA-1(ch)   KSQD-----DVIPPAKNLRDAGIELFAIGVKNAD-----------INELKEIASE-PDSTH
collXII VA-4(h)    RSQD-----EVKKAALVIQQSGFSVFVVGVADVD-----------YNELANIASK-PSERH
collVII VA(h)      KSQD-----LVDTAAQRLKGQGVKLFAVGIKNAD-----------PEELKRVASQ-PTSDF
collVI VA-1(m)     QSED-----GFALPSAELKSADVNVFAVGVEGAD-----------ERALGEVASE-PLSMH
WARP VA(m)         GSSD-----PVGPPMQELKDLGVTIFIVSTGRGN-----------LLELLAAASA-PAEKH
cochlin VA-1(ch)   WPSD-----DLEEAGIVAREFGVNVFIVSVAKPT-----------TEELGMVQDIGFIDKA
VLA VA(h)          ESHD---NHRLKKVIQDCEDENIQRFSIAILGSYNRGNLSTEKFVEEIKSIASE-PTEKH
vWF VA-A1(h)       QEPQRMS-RNFVRYVQGLKKKKVIVIPVGIGPHAN----------LKQIRLIEKQAP-ENK
```

Figure 2A

```
mat2 VA-1(m)          VFLVANFSQIESLTSVFQNKLCTV
mat4 VA-1(m)          VFLVESFDLIQEFGLQFQGRLCGK
mat3 VA-1(m)          VFYVETYGVIEKLSARFQETFCAL
mat1 VA-1(m)          VDYVESYNVIEKLAKKFQEAFCVV
collXIV VA-1(ch)      VYNVADFNFMNSIVEGLTRTVCSR
collXII VA-4(h)       VFIVDDFESFEKIEDNLITFVCET
collVII VA(h)         FFFVNDFSILRTLLPLVSRRVCTT
collVI VA-1(m)        VFNLENVTSLHGLVGNLVSCIHSS
WARP VA(m)            LHFVD-VDDLPIIARELRGSITDA
cochlin VA-1(ch)      VCRNNGFFSYQMPSWFGTTKYVKP
VLA VA(h)             FFNVSDELALVTIVKTLGERIFAL
vWF VA-A1(h)          AFVLSSVDELEQQRDEIVSYLCDL
```

Figure 2A continued

```
                   A             B                     C                      C'
WARP F3-2(m)       PERIVISHARPRSLRVSWAPALGPDSALGYHVQLGPLQG-GS---LER--VEVPAGQNS-
colVII F3-1(m)     PRDLVLSEPSSQSLRVQWTAASGP--VTGYKVQYTPLTGLGQPLPSERQEVNIPAGETS-
colXIV F3-5(ch)    PQHLEVDEASTDSFRVSWKPT-SSDIAF-YRLAWIPLDG-GE---SEE--VVLSGDADS-
 4int F3-3(h)      PTRLVFSALGPTSLRVSWQEPRCERPLQGYSVEYQLLNG-GE---LHR--LNIPNPAQTS
colXII F3-3(m)     PRNLKVTDETTDSFKLTWSQAPGR--VLRYRIRYRPVSG-GE---SKE--VSTPANQRR-
fibro F3-12(m)     PSQMQVTDVQDNSISVRWLPSTSP--VTGYRVTTTPKNGLGP----SK--TKTASPDQTE
WARP F3-1(m)       PQQLHASEVLSSGFRLSWPPLLTAD-SGYYVLELVPSGKLAT---TRR--QQLPGNATS-
tenascR F3-7(h)    PKDITTISNVTKDSVMVSWSPPVAS--FDYYRVSYRPTQV-GR---LDS--SVVPNTVTE-

                   E         F              G
WARP F3-2(m)       TTVQGLTPCTTYLVIVTAAFRSGRQRALSAKACT
colVII F3-1(m)     TRLQGLRPLTDYQVIVVALYANSIGEAVSGTART
colXIV F3-5(ch)    YVIEGLLPNTEYEVSLLAVEDDETESEVVAVLGA
 4int F3-3(h)      VVVEDLLPNHSYVFRVRAQSQEGWGREREGVITI
colXII F3-3(m)     KTLENLTPDTKYEISVIAEYPSGPGSPLTGNAAT
fibro F3-12(m)     MTIEGLQPTVEYVVSVYAQNRNGESQPLVQTAVT
WARP F3-1(m)       WTWTDLDPDTDYEVSLLPESNVHLLRPQHVRVRT
tenascR F3-7(h)    FTITRLNPATEYEISLNSVRGREESERICTLVHT
```

Figure 2B

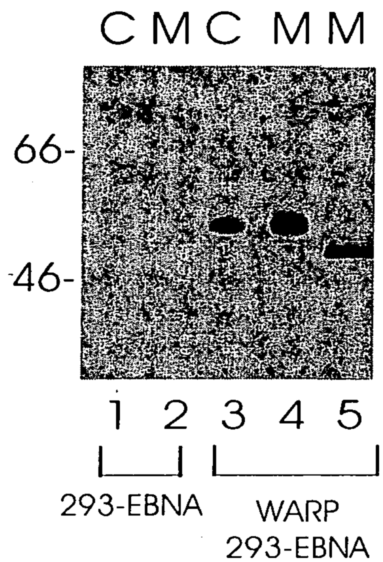


Figure 3

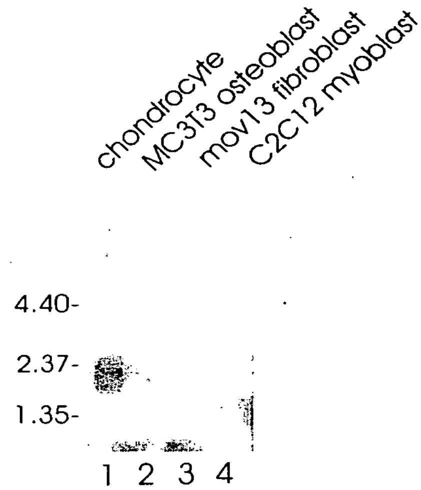


Figure 4A

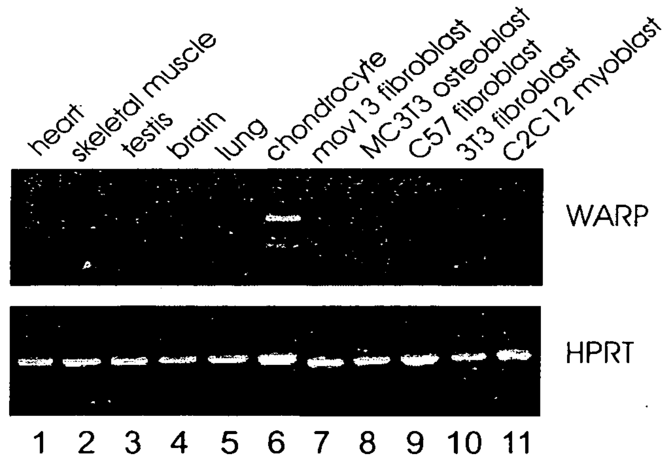
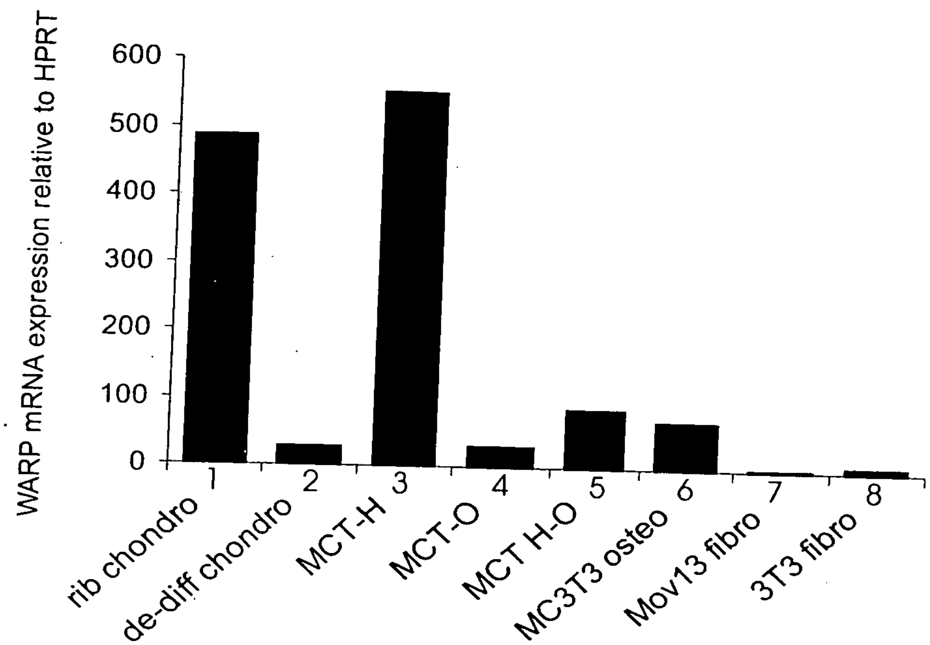


Figure 4B



WARP mRNA expression relative to HPRT
600
500
400
300
200
100
0
1
2
3
4
5
6
7
8
rib chondro
de-diff chondro
MCT-H
MCT-O
MCT H-O
MC3T3 osteo
Mov13 fibro
3T3 fibro

Figure 4C

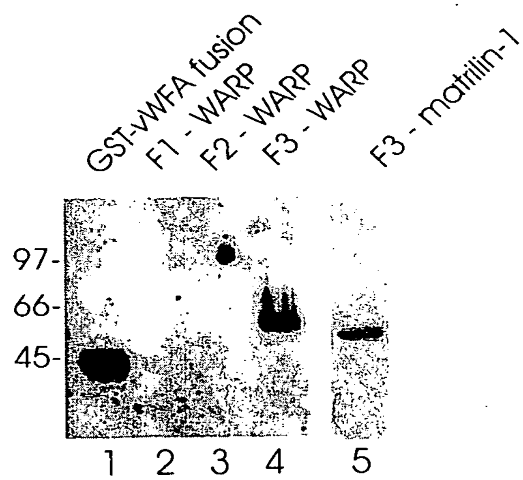


Figure 5A

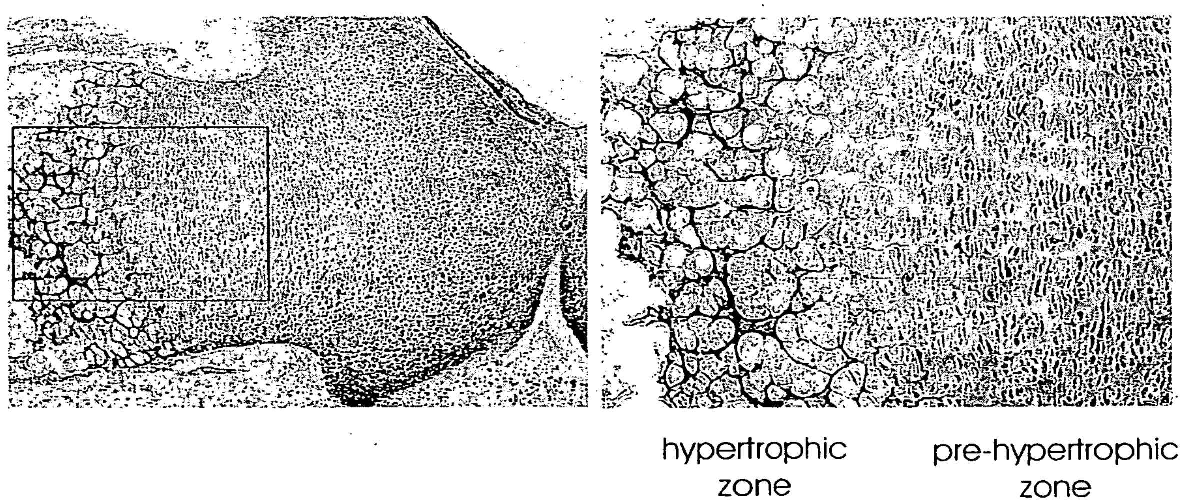


Figure 5B

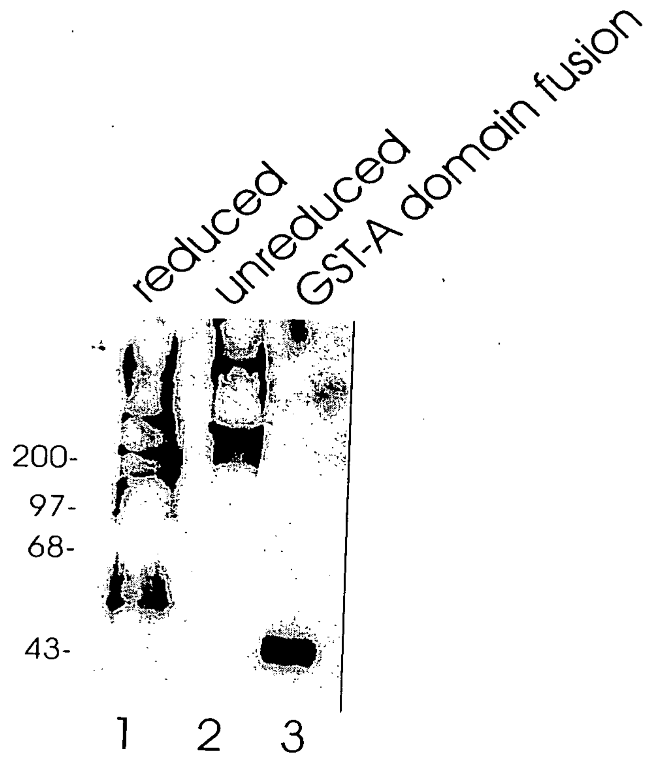


Figure 6